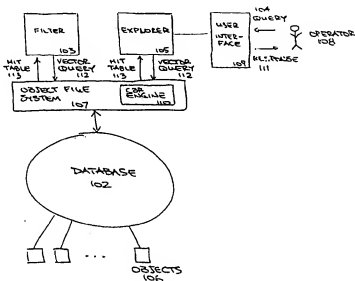




## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>5</sup> : <b>G06F 15/40</b>	<b>A1</b>	(11) International Publication Number: <b>WO 95/02221</b>
		(43) International Publication Date: 19 January 1995 (19.01.95)
(21) International Application Number: <b>PCT/US94/07569</b>	(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, ES, FI, GB, GE, HU, JP, KE, KG, KP, KR, KZ, LK, LT, LU, LV, MD, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SI, SK, TJ, TT, UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CG, CF, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD).	
(22) International Filing Date: 5 July 1994 (05.07.94)		
(30) Priority Data: 08/088,307 2 July 1993 (07.07.93) US		
(71) Applicant (for all designated States except US): INFERRECE CORPORATION [US/US]; 550 N. Continental Boulevard, El Segundo, CA 90245 (US).	Published With international search report.	
(72) Inventors; and		
(75) Inventors/Applicants (for US only): ALLEN, Bradley, P. [US/US]; 829 Loma Drive, Hermosa Beach, CA 90254 (US). LEE, David, J. [US/US]; 5523 161st Place, N.E., Redmond, WA 98052 (US). CARASSO, Roger, D. [US/US]; Suite 430, 640 S. Masselin Avenue, Los Angeles, CA 90036 (US). PERRY, John, R. [US/US]; 6082 Dagny Circle, Huntington Beach, CA 92647 (US).		
(74) Agents: SWERNOFSKY, Steven, A. et al.; D'Alessandro, Frazzini & Ritchie, Suite 101 & 201, 2099 Lincoln Avenue, San Jose, CA 95125 (US).		

## (54) Title: CASE-BASED ORGANIZING AND QUERYING OF A DATABASE



## (57) Abstract

A system for case-based organizing and querying of a database (102). The database (102) may comprise a set of objects (106), such as text documents. The database (102) may be organized by examining each object (106) and associating that object (106) with a set of property values, such as keywords. A document may be associated with those words which appear more frequently in the document than in the database (102) at large, or which appear in the early text of the document, or which appear in the title. The system may be responsive to a query (104) by associating the query with a similar set of property values and performing case-based matching on the objects (106) of the database (102) for similar objects (106). The query (104) may be natural-language text and may be associated with keywords. The system may present matched objects in response to the query (104), may respond to iterative refinement of the query and may order matched objects by quality of match. The system may also respond to the result of organizing matched objects for presentation with suggestions for iterative refinement of the query (104).

# FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CZ	Czechoslovakia	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
ES	Spain	MG	Madagascar	UA	Ukraine
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

## CASE-BASED ORGANIZING AND QUERYING OF A DATABASE

1. Field of the Invention

This invention relates to case-based organizing and querying of a database.

2. Description of Related Art

As storage capability grows for computing devices, many databases have become larger, and large databases have become more common. One problem which has become apparent in the art is the difficulty of retrieving information from large databases when the location of that desired information is not already known. For example, a search for information in a large library may be hampered by the size of the library, because of the large number of items which must be examined. This can be exacerbated if the information searched for is not well-described by the searcher, if the searcher is unfamiliar with that subject matter, or if the information searched for is not well indexed.

Large databases of objects may sometimes be generated without the original intent to organize them into a database. For example, newspaper articles may generally be written without the consideration that they may be collected into a single database for later search. When they eventually are collected into a database, the effort required to organize those objects into a database for information retrieval can be formidable. It

1 would be advantageous to provide a system in which a large amount  
2 of information may be collected into a database without having to  
3 expend a comparable amount of effort on organization and  
4 indexing, e.g., where such organization and indexing can be done  
5 by an automated process.

6  
7           Prior art methods of retrieving information generally  
8 require preparation of a query, in which objects to be searched  
9 for are described in some formal manner. This imposes additional  
10 effort on the searcher, and generally also requires that the  
11 searcher be familiar with the subject matter to be searched, with  
12 the organization and indexing of the database, and with a formal  
13 query language. Accordingly, it would be advantageous for the  
14 searcher to be able to describe the query in a natural and  
15 relatively informal or unstructured manner, such as a description  
16 in a natural language.

17  
18           Work with case-based systems has shown that incremental  
19 refinement of problem descriptions can be valuable in improving a  
20 automated system's recall (ability to retrieve objects which are  
21 related to the query) and precision (ability to rule out objects  
22 which are not related to the query). It would be advantageous to  
23 be able to incrementally refine the query after a response. But  
24 when the query itself is unstructured, the original response may  
25 provide so much information that valuable material is lost in the  
26 size of the response. Accordingly, it would be advantageous to  
27 provide suggestions for incremental refinement. In one aspect of  
28 the invention, the response may be organized by quality of match.

1 In another aspect, the response may be organized into clusters of  
2 related objects.

3  
4 SUMMARY OF THE INVENTION

5  
6 The invention provides a system for case-based  
7 organizing and querying of a database. The database may comprise  
8 a set of objects, such as a set of documents including text. In  
9 a preferred embodiment, the database may be organized by  
10 examining each object and associating that object with a set of  
11 property values, such as (in the case of text documents) a set of  
12 keywords or other indicators of content. For example, a document  
13 may be associated with those words which appear more frequently  
14 in the document than in the database at large, or which appear in  
15 early text of the document, or which appear in a title. The  
16 system may be responsive to a query by associating the query with  
17 a similar set of property values and performing case-based  
18 matching or other fuzzy associative matching on the objects of  
19 the database for objects which are similar. In a preferred  
20 embodiment, the query may be natural-language text and may be  
21 associated with keywords or other indicators of its content.

22  
23 In a preferred embodiment, the system may present  
24 matched objects in response to the query, may respond to  
25 iterative refinement of the query (in similar manner to iterative  
26 case-based methods shown in those co-pending applications which  
27 have been incorporated by reference), and may order matched  
28 objects by quality of match. The system may also examine the

1 collection of matched objects and organize them for presentation;  
2 for example, the system may group matched objects into clusters  
3 of objects which have similar properties, which relate to similar  
4 content, or which have similar likelihood to be of relevance to  
5 the query or of interest to an operator posing the query. The  
6 system may respond to the result of organizing matched objects  
7 for presentation with suggestions for iterative refinement of the  
8 query.

9  
10 The system may therefore be capable of producing  
11 improved recall and precision over prior art techniques.

#### 12 BRIEF DESCRIPTION OF THE DRAWINGS

13  
14  
15 Figure 1 shows a block diagram of a database explorer  
16 and filter system.

17  
18 Figure 2 shows a data flow diagram of a method of  
19 filtering documents.

20  
21 Figure 3 shows a data flow diagram of a method of  
22 processing queries.

23  
24 Figure 4 shows a data flow diagram of a method of  
25 processing hit tables.

26  
27 Figure 5 shows a process flow diagram of a method of  
28 clustering hit tables.

1           Figure 6 shows an example explorer user interface  
2 screen as viewed by an operator.

3  
4           Figure 7 shows a second example explorer user interface  
5 screen, as viewed by an operator, in which clusters are  
6 displayed.

7  
8           Figure 8 shows an example explorer user interface  
9 screen, as viewed by an operator, in which settings may be set by  
10 the operator.

11  
12           Appendix A shows a table of parts of speech and a set  
13 of lexical rules for the English language, which may be used for  
14 the tag-and-segment-text process or the tag-and-segment-text  
15 process in a preferred embodiment.

16  
17           Appendix B shows an output of a test run of an example  
18 filter when applied to a portion of an example multimedia  
19 encyclopedia used as a database, available as "Microsoft Encarta"  
20 from Microsoft Corporation of Redmond, Washington.

#### 21 22           DESCRIPTION OF THE PREFERRED EMBODIMENT

23  
24           An embodiment of this invention may be used together  
25 with inventions which are disclosed in a copending application  
26 titled "AUTONOMOUS LEARNING AND REASONING AGENT", application  
27 Serial No. 07/ 869,926, filed April 15, 1992 in the name of  
28

1 Bradley P. Allen, hereby incorporated by reference as if fully  
2 set forth herein.

3  
4 In a preferred embodiment, the invention may operate in  
5 conjunction with a computing system, including a processor and a  
6 memory, generally configured as is well known in the art; the  
7 memory may include primary memory for stored programs and for  
8 data and secondary memory for extensive storage of large numbers  
9 of objects. Preferably, the memory may comprise a sizable  
10 database of objects, as is well known in the art of databases,  
11 and such objects may comprise various types of computing and  
12 data-storage structures. However, no particular structure is  
13 required for the database itself; the database may be a  
14 relational database, an unstructured collection of objects, or  
15 some other database format.

16  
17 Although the invention is disclosed herein primarily  
18 with respect to textual objects, it would be clear to those of  
19 ordinary skill in the art, after perusal of the application, that  
20 extension of the concepts disclosed to other types of objects is  
21 within the scope and spirit of the invention, and would not  
22 require undue experimentation. Such other types of objects may  
23 include source code, object code, binary values, numeric values,  
24 text or other symbolic values, representations of sound and/or  
25 picture signals or other signals, multimedia, data structures for  
26 rule-based or case-based systems, artificial neural networks,  
27 linked data structures such as linked lists, mathematical  
28 structures such as equations, polynomials, matrices or tensors,



1 and other data types known in at least one of the many fields of  
2 computing. Although when the invention is applied to textual  
3 objects, appearance of a text string in an object is considered  
4 pertinent, when the invention is applied to other types of  
5 objects, other measures of closeness or pertinence, such as  
6 numerical closeness, would be workable, and are within the scope  
7 and spirit of the invention.

#### 8 9 FILTER AND EXPLORER SYSTEM

10  
11 Figure 1 shows a block diagram of a database explorer  
12 and filter system.

13  
14 In a preferred embodiment, a system 101 for case-based  
15 organizing and querying of a database 102 may comprise a filter  
16 103, for organizing the database 102 so as to be responsive to a  
17 query 104, an explorer 105, for selecting a set of objects 106 in  
18 the database 102 which are responsive to that query 104, and an  
19 object file system 107, for accessing the database 102. In a  
20 preferred embodiment, the database 102 may generally be of a type  
21 which is known in the art, such as a collection of text objects  
22 supported by Cairo Milestone 4 running under the Windows NT  
23 system version 297, available from Microsoft Corporation of  
24 Redmond, Washington, and may be accessed in conjunction with the  
25 object file system 107 of that product.

26  
27 The filter 103 may operate at an initialization time,  
28 such as when the processor is first started or before the first

1 query 104 is presented to the explorer 105. The filter 103 may  
2 also operate in an incremental mode, e.g., by updating its  
3 organization of the database 102 periodically, such as upon the  
4 passage of a fixed period of time, when a fixed number of objects  
5 106 are changed or added to the database 102, when the operation  
6 of the explorer 105 is degraded below some predetermined level,  
7 when triggered by an operator 108 in conjunction with a user  
8 interface 109 (e.g., when a query is presented, by a specific  
9 command to do so, or as a side effect of another operation), or  
10 otherwise as determined by the database 102 or an external  
11 manager.

12  
13           The filter 103 may examine each of the objects 106 (or  
14 some predetermined subset of objects 106) in the database 102 and  
15 associate each object 106 it examines (or some predetermined  
16 subset of those objects 106) with a set of properties. For a  
17 textual database 102 as primarily described herein, those  
18 properties may be keywords or phrases which are found in the  
19 object 106, but may also comprise other property values, such as  
20 the language the text is written in, the length of the text, or  
21 the reading level or other measure associated with the text  
22 (including measures of complexity, detail, redundancy, writing  
23 style, "fog", or other known measures of text, e.g., known in the  
24 art of grammar checking and correction).

25  
26           The objects 106 with their properties may be treated as  
27 a set of cases to be matched by a CBR engine 110 (operating with  
28 the object file system 107) with a test case generated from the

1 query 104. Each case may generally comprise an object 106 plus  
2 the properties that object 106 was associated with, e.g., key  
3 words and phrases found in that object. In a preferred  
4 embodiment, these properties may include a lexicon of words and  
5 noun phrases found in the object 106, including at least some of  
6 these words labelled as a set of "header words" or "relevant  
7 words".

8  
9           The explorer 105 may generally operate at a question  
10 time, such as when one or more queries 104 is presented to the  
11 explorer 105. In a preferred embodiment, the explorer 105 may be  
12 invoked by the operator 108 in conjunction with the user  
13 interface 109, which user interface 109 may allow the operator to  
14 trigger operation of the explorer 105 and to present one or more  
15 queries 104 to the explorer 105. In a preferred embodiment, the  
16 user interface 109 may be one such as the user interface  
17 presented by the Windows NT system referred to herein. In a  
18 preferred embodiment, the operator 108 may be a human being, but  
19 those of ordinary skill with recognize, after perusal of the  
20 application, that the operator 108 may comprise a network  
21 connection, an external management program, or an AI program.

22  
23           In a preferred embodiment, the explorer 105 may  
24 generate a response 111 including a set of matching cases (i.e.,  
25 objects 106 with their properties), which may be presented to the  
26 operator 108 by means of the user interface 109, such as the user  
27 interface presented by the Windows NT system referred to herein,  
28 augmented by features described herein.

The filter 103 and the explorer 105 may operate in conjunction with the object file system 107 (and in particular the CBR engine 110 thereof), which may respond to a set of properties formed into a vector query 112 directed at the database 102, and may return a hit table 113 of those objects 106 in the database 102 which have the indicated properties. In a preferred embodiment, the CBR engine 110 may use case-based matching and other techniques such as those shown in those co-pending applications which have been incorporated by reference.

## FILTERING DOCUMENTS

Figure 2 shows a data flow diagram of a method of filtering documents.

In a preferred embodiment, a document 201 (an object 106 which comprises text, such as a pure text document or a text document formatted for a word-processing program) may be input to the filter 103 for examination. The filter 103 may process the text by a tag-and-segment-text process 202, which may lexically analyze the document 201, e.g., by means of a known lexical analysis technique.

The tag-and-segment-text process 202 may extract a set of single terms 203 and generate a set of header words 204 found in the document 201. The header words 204 may comprise those words which occur in an initial part of the object 106, or in a title, subject line, topical paragraph, or abstract. In a

1 preferred embodiment, the header words 204 may comprise the first  
2 three things mentioned in the document 201.

3  
4         The tag-and-segment-text process 202 may also tag words  
5 in the document 201 with their parts of speech and parse them  
6 into a set of sentences 205. The sentences 205 may be input to  
7 an extract-noun-phrases process 206, which may further lexically  
8 analyze the document 201, e.g., by means of a known lexical  
9 analysis technique, to extract a set of noun phrases 207 and  
10 generate a lexicon 208 thereof. In a preferred embodiment, the  
11 tag-and-segment-text process 202 may use a grammar of the English  
12 language, but other natural languages, and even formal  
13 specification languages such as programming languages, would also  
14 be suitable.

15  
16         The tag-and-segment-text process 202 may also recognize  
17 and generate a set of proper nouns 209. In a preferred  
18 embodiment, the set of proper nouns 209 may be determined by  
19 known rules, e.g., that proper nouns generally comprise strings  
20 of words each starting with an upper-case letter, or by reference  
21 to a dictionary of known proper names. The set of proper nouns  
22 209 may be input, along with at least some of the single terms  
23 203, to a determine-relevant-words process 210, which may extract  
24 a set of relevant words 211.

25  
26         The set of relevant words 211 may be determined with  
27 reference to the frequency of those words in the object 106 (with  
28 respect to the entire text found in the object 106) and with

1 reference to the frequency of those words in the database 102,  
2 with respect to the text corpus of the database 102. In a  
3 preferred embodiment, the ratio for each word (frequency in the  
4 object 106) divided by (frequency in the database 102) may be  
5 computed, and the set of relevant words 211 may comprise those  
6 words whose relative frequency exceeds a threshold, e.g., a  
7 predetermined threshold such as a 1:1 ratio. However, it would  
8 be clear to those of ordinary skill, after perusal of this  
9 application, that other measures (e.g., statistical measures)  
10 relating to frequency could be used to determine relevant words,  
11 such as clustering of relevant words in paragraphs, correlation  
12 with other relevant words, or relative frequency of word pairs or  
13 n-tuples, and that such other measures are within the scope and  
14 spirit of the invention.

15

16           The filter 103 is described herein for a specific set  
17 of properties of the text which may be extracted. However, it  
18 would be clear to those of ordinary skill, after perusal of this  
19 application, that extraction of other properties could be readily  
20 accomplished, and is within the scope and spirit of the  
21 invention. Such other properties could include the language the  
22 text is written in (or for English-language text, the number of  
23 foreign words used), the length of the text, or the reading level  
24 or other measure associated with the text (including measures of  
25 complexity, detail, redundancy, writing style, "fog", or other  
26 known measures of text, e.g., known in the art of grammar  
27 checking and correction).

28

1 In a preferred embodiment, the extract-noun-phrases  
2 process 206 and the determine-relevant-words process 211 may  
3 proceed in parallel, e.g., by execution on multiple processors or  
4 by multiple tasks or threads in a multitasking or multithreaded  
5 environment.

6  
7 The filter 103 may mark each object 106 with the  
8 properties it determines (or alternatively may create a separate  
9 object 106 relating each documentary object 106 to its  
10 properties), so that the object 106 and its properties may be  
11 treated as a case in a case-base. In a preferred embodiment, the  
12 set of cases may be matched to a test case by a CBR engine 110,  
13 using techniques like those described in copending applications  
14 (1) Serial No. 07/ 664,561, filed March 4, 1991 in the name of  
15 inventors **Bradley P. Allen** and **S. Daniel Lee**, titled "CASE-BASED  
16 REASONING SYSTEM"; (2) Serial No. 07/ 869,935, filed April 15,  
17 1992 in the name of inventor **Bradley P. Allen**, titled "MACHINE  
18 LEARNING WITH A RELATIONAL DATABASE"; and (3) Serial No. 07/  
19 869,926, filed April 15, 1992 in the name of **Bradley P. Allen**,  
20 titled "AUTONOMOUS LEARNING AND REASONING AGENT"; each of which  
21 is hereby incorporated by reference as if fully set forth herein,  
22 or other case-based reasoning techniques which may be known in  
23 the art.

#### 24 PROCESSING QUERIES

25  
26  
27 Figure 3 shows a data flow diagram of a method of  
28 processing queries.

1 In a preferred embodiment, the query 104, entered in  
2 free text by the operator 108, may be input to the explorer 105  
3 for examination. The explorer 105 may process the text by a tag-  
4 and-segment-text process 301, which may lexically analyze the  
5 document 201, e.g., by means of a known lexical analysis  
6 technique, similarly to the tag-and-segment-text process 202 of  
7 the filter 103.

8  
9 The tag-and-segment-text process 301 may extract a set  
10 of single terms 302, similarly to the tag-and-segment-text  
11 process 202 and the set of single terms 203 of the filter 103.

12  
13 The tag-and-segment-text process 301 may also tag words  
14 in the document 201 with their parts of speech and parse them  
15 into a set of sentences 303, similarly to the tag-and-segment-  
16 text process 202 and the sentences 205 of the filter 103. The  
17 sentences 303 may be input to an extract-noun-phrases process  
18 304, which may further lexically analyze the document 201, e.g.,  
19 by means of a known lexical analysis technique, to extract a set  
20 of noun phrases 305, similarly to the extract-noun-phrases  
21 process 206 and the noun phrases 207 of the filter 103.

22  
23 The tag-and-segment-text process 301 may also recognize  
24 and generate a set of proper nouns 306, similarly to the tag-and-  
25 segment-text process 202 and the proper nouns 209 of the filter  
26 103.



The noun phrases 305, single terms 302, and proper nouns 306, a rank threshold 307, and a set of selected subtopics 308 (subtopics selected by the operator 108 to refine the query 104) may be input to a generate-query process 309, which may generate a set of query terms 310 and a query parse tree 311.

In a preferred embodiment, the tag-and-segment-text process 301, the extract-noun-phrases process 304, and the generate-query process 309 may proceed as asynchronously as possible, e.g., by execution on multiple processors or by multiple tasks or threads in a multitasking or multithreaded environment.

The query terms 310 and the query parse tree 311 may be input to the CBR engine 110 in the object file system 107, and may perform case-based matching or other fuzzy associative matching on the objects 106 in the database 102 for objects which are similar to the query 104, as described by the query terms 310 and the query parse tree 311, and which have a match quality at least as good as the rank threshold 307. (As noted with regard to the user interface 109, the selected subtopics 308 are added to the text of the query 104.) The object file system 107 may generate the hit table 113 of matched objects 106.

## PROCESSING HIT TABLES

Figure 4 shows a data flow diagram of a method of processing hit tables.

1           The hit table 113 and the relevant words 211 may be  
2 input to a cluster hits process 401, which (if clustering is  
3 enabled) collects the matched objects 106 into clusters, and may  
4 output a set of clusters 402 in response. Each cluster 402 may  
5 comprise a set of objects 106, selected for collective closeness  
6 with regard to all objects 106 in the hit table 113. The cluster  
7 hits process 401 is further described with regard to figure 5.

8  
9           The hit table 113, the relevant words 211, and the  
10 lexicon 208 may be input to a first generate-topics (from  
11 relevant words) process 403, while the lexicon 208 and the query  
12 terms 310 may be input to a second generate-topics (from query  
13 words) process 403. Together the two generate-topics processes  
14 403 may output a set of topics 404 and subtopics 405.

15  
16           In a preferred embodiment, the generate-topics process  
17 403 may examine the lexicon 208 of noun phrases 207 with a rule-  
18 based inference engine (not shown). (One such inference engine  
19 is the ART-IM system, available from Inference Corporation in El  
20 Segundo, California.) The inference engine may detect particular  
21 patterns in the noun phrases 207 which indicate semantic  
22 relations between the words in those noun phrases 207. For  
23 example, the noun phrase

24  
25           "kangaroos, wallabies, and other marsupials"  
26  
27 would be detected and would generate the relations  
28

1 kangaroo IS-A marsupial

2 wallaby IS-A marsupial

3  
4 The generate-topics process 403 may thus construct a  
5 phrase lattice, showing each noun phrase 207 as being inclusive  
6 of (above), included in (below), or incommensurate with (neither  
7 above nor below) each other noun phrase 207.

8  
9 The generate-topics (from relevant words) process 403  
10 may restrict the phrase lattice to those noun phrases 207 which  
11 include relevant words 211 of the objects 106 in the hit table  
12 113. In a preferred embodiment, the second generate-topics (from  
13 query words) process 403 may operate in similar manner as the  
14 first generate-topics (from relevant words) process 403 and may  
15 restrict the phrase lattice to those noun phrases 305 which  
16 include relevant words 211 of the query.

17  
18 Figure 5 shows a process flow diagram of a method of  
19 clustering hit tables.

20  
21 The cluster hits process 401 may operate by means of a  
22 genetic algorithm, in which an initial configuration and a set of  
23 genetic operators are specified, and the set of solutions is  
24 formed by simulation of random "evolution" of a population of  
25 possible solutions, using the method of steady-state reproduction  
26 without duplicates. Genetic algorithms are well known in the  
27 art, and are described in further detail in "Foundations of  
28 Genetic Algorithms", ed. Gregory J.E. Rawlins (Morgan Kaufmann

1 Publishers: San Mateo, California 1991). It would be clear to  
2 those of ordinary skill in the art that the parameters of the  
3 genetic algorithm, and even the type of genetic algorithm  
4 performed could be varied substantially and still remain within  
5 the scope and spirit of the invention.

6  
7 In a cluster-count step 501, a number of clusters 402  
8 is selected. The number of clusters 402 may vary from a known  
9 minimum to a known maximum, settable by the operator 108. The  
10 genetic algorithm of the following steps is repeated for each  
11 permissible number of clusters 402, and the best solution  
12 adopted.

13  
14 In an initiate-clusters step 502, a set of possible  
15 clusters 402 is selected; this is a single "gene". A random  
16 population of genes is selected. Each cluster 402 is represented  
17 by the centroid of the objects 106 which would comprise that  
18 cluster 402. Thus, when a solution of clusters 402 is selected,  
19 each object 106 is assigned to the cluster 402 which it best  
20 matches.

21  
22 After the initiate-clusters step 502, the genetic  
23 algorithm of the following steps is repeated for a known period  
24 of time, settable by the operator 108. When that time expires,  
25 the best available solution (i.e., the gene with the best  
26 quality) is selected as the solution and specifies the set of  
27 clusters 402. Each object 106 is assigned to the cluster 402 to  
28 which it is the closest.

1           In an evaluation step 503, all genes in the population  
2 are evaluated for quality, and the gene with the least quality is  
3 removed. In a preferred embodiment, the statistical measure  
4 "category utility" is computed; i.e., the utility of each cluster  
5 402 in distinguishing between an object 106 in one cluster 402  
6 from an object in another cluster 402. Thus, if the centroid of  
7 a cluster 402 has high quality of match for several objects 106,  
8 those objects are reasonably clustered together.

9  
10           Although in a preferred embodiment, matching for  
11 clusters 402 is performed using relevant words 211, it would be  
12 clear to those of ordinary skill, after perusal of this  
13 application, that other properties of the objects 106 could be  
14 used as well, such as the read/write date of the object 106, and  
15 that doing so would be within the scope and spirit of the  
16 invention.

17  
18           In a genetic-operator step 504, one of three operators  
19 is selected and employed to create a new gene: (1) Mutation-1.  
20 The new gene is randomly created. (2) Mutation-2. An existing  
21 gene is copied, except that one of its clusters 402 is mutated by  
22 replacing it with a randomly created cluster 402. (3) Crossover.  
23 Two genes have their n-tuples of clusters 402 paired off and one  
24 cluster 402 is selected at random from each pair to form the new  
25 gene. Alternatively, a new gene is created by selecting N  
26 clusters 402 at random from the 2N clusters 402 specified by the  
27 two old genes.

## USER INTERFACE

Figure 6 shows an example explorer user interface screen as viewed by an operator. While the invention is described primarily with regard to a specific user interface, it would be clear to those of ordinary skill in the art that another user interface of equal or greater flexibility would be suitable, and would be within the scope and spirit of the invention.

In a preferred embodiment, the user interface 109 may be combined with a user interface for a generalized file system exploration program, such as in the Windows NT system referred to herein. The user interface 109 may comprise a query window 601 in which the operator may enter the query 104 in free text, and a results window 602 in which the system 101 may display a set of matched objects 106 found in response to the query 104.

In a preferred embodiment, the operator 108 may enter the query 104 in the query window 601. The query 104 is input to the explorer 105, which processes it as described herein, and generates the vector query 112. The vector query 112 is input to the object file system 107, and generates the hit table 113 of matched objects 106. The hit table 113 is input to the user interface 109, which displays the matched objects 106. The operator may select a displayed matched object 106 to view its contents.

1           In a preferred embodiment, the user interface 109, the  
2 explorer 105, and the object file system 107, may operate as  
3 asynchronously as possible. Accordingly, the object file system  
4 107 may search the database 102 for matched objects 106  
5 independently, once it has sufficient information from the  
6 explorer 105; the user interface 109 may display matched objects  
7 106 from the hit table 113 as they are generated by the object  
8 file system 107.

9  
10           In the example, the operator 108 has entered the query  
11 104 "who invented the light bulb?" in a content field 603 of the  
12 query window 601, and the system 101 has responded with a set of  
13 matched objects 106 in the results window 602. The matched  
14 objects are displayed one per line, in columns labelled "rank",  
15 "query", "header", and "relevant words".

16  
17           In the example, a rank field 604 displays the quality  
18 of match for each displayed matched object 106. In a preferred  
19 embodiment, the system 101 may order the matched objects 106 by  
20 rank. This may occur as the normal procedure, or at the request  
21 of the operator 108, e.g., by means of a "sort" command 605 in  
22 the query window 601. In a preferred embodiment, the rank field  
23 604 may also be color-coded by value.

24  
25           In the example, a query field 606 displays the relevant  
26 words of the query which are most related to the displayed  
27 matched object 106.

In the example, a header field 607 displays the header words 204 of the displayed matched object 106.

In the example, a relevant words field 608 displays the most common relevant words 211 of the displayed matched object 106.

In the example, a topics field 609 of the query window 601 displays suggested topics for refinement of the query 104 which the system 101 has identified. In a preferred embodiment, the operator 108 may select a topic in the topics field 609, and the system will display a subtopics window 610 (overlaid on the query window 601 and the results window 602) showing the subtopics which the system 101 has identified for that topic.

## QUERY REFINEMENT

The operator 108 may refine the query 104 in response to the matched objects 106, and the explorer 105 may attempt to match objects 106 using the query 104 as refined. This may occur at the request of the operator 108, e.g., by means of a "refresh" command 611 in the query window 601.

In a preferred embodiment, the operator 108 may select one or more subtopics 405 to refine the query 104. To do so, the operator 108 may identify (e.g., by pointing to with a pointing device such as a mouse) one or more subtopics 405 in the subtopics window 610. The selected subtopics 308 may be "added"



1 to the query 104 and the explorer 105 may attempt to match  
2 objects 106 using the query 104 as refined.

3  
4 In a preferred embodiment, the operator 108 may also  
5 select one or more relevant words 211 to refine the query 104. To  
6 do so, the operator 108 may identify (e.g. by pointing to) the  
7 relevant words field 608 for a particular matched object 106 and  
8 "drag" that relevant words field 608 to the content field 603;  
9 the system 101 will display a relevance feedback window 612  
10 (overlaid on the query window 601 and the results window 602)  
11 showing the relevant words 211 for that matched object 106.

12  
13 In a preferred embodiment, the operator 108 may select  
14 one or more relevant words 211 to refine the query 104. To do  
15 so, the operator 108 may identify (e.g., by pointing to) one or  
16 more relevant words 211 in the relevance feedback window 612.  
17 The selected relevant words 211 may be "added" to the query 104  
18 and the explorer 105 may attempt to match objects 106 using the  
19 query 104 as refined.

20  
21 The query 104 as refined (like the original query 104)  
22 is presented as a vector query 104 to the CBR engine 110. When  
23 selected subtopics 308 or relevant words 211 are "added" to the  
24 query, they are properties which the CBR engine 110 must match to  
25 objects 106, as described for methods of iterative refinement of  
26 case-based matching shown in those co-pending applications which  
27 have been incorporated by reference. (Thus, the CBR engine 110  
28 must match to objects 106 as if the operator 108 had answered a

1 query refining question in a case-based system.) A query 104 as  
2 refined may be further refined, allowing the operator to  
3 iteratively refine the query 104 until desired objects 106 are  
4 located.

#### 6 VIEWING CLUSTERS

8 Figure 7 shows a second example explorer user interface  
9 screen, as viewed by an operator, in which clusters are  
10 displayed.

12 The operator 108 may select a "cluster" command (figure  
13 6) or "uncluster" (figure 7) command 701 in the query window 601,  
14 and the system 101 will display a set of clusters 402, each a set  
15 of related matched objects 106, in place of displaying matched  
16 objects 106 themselves. In the example, the operator has  
17 selected the "cluster" command 701 for the same query 104 as in  
18 the example of figure 6.

20 In the example, an expand field 702 displays whether  
21 the cluster 402 can be expanded (shown by a "+" symbol) to  
22 display individual matched objects 106, or can be collapsed  
23 (shown by a "-" symbol) to display a single identifier for the  
24 cluster 402.

26 In the example, the rank field 703 displays the best  
27 rank for all matched objects 106 in the cluster 402. In a  
28 preferred embodiment, the system 101 may order the clusters 402

1 by this rank field 703. This may occur as the normal procedure,  
2 or at the request of the operator 108, e.g., by means of the  
3 "sort" command 605 in the query window 601. In a preferred  
4 embodiment, this rank field 703 may also be color-coded by value.

5  
6 In the example, the relevant words field 608 displays  
7 the most common relevant words 211 in the cluster 402.

8  
9 Other fields and windows remain similar to the example  
10 of figure 6.

11  
12 The operator 108 may also choose to cluster all objects  
13 106 in a specific set, e.g., a specific directory in the object  
14 file system 107. In a preferred embodiment, the operator 108 may  
15 restrict the scope of the explorer 105 to a specific directory  
16 and issue the "cluster" command 701; the system 101 will display  
17 the objects 106 in that directory in clusters 402.

#### 18 19 SETTING PARAMETERS

20  
21 Figure 8 shows an example explorer user interface  
22 screen, as viewed by an operator, in which settings may be set by  
23 the operator.

24  
25 In a preferred embodiment, the operator 108 may select  
26 settings appropriate for the system 101. The operator 108 may  
27 select a "properties" command 801 in the query window 601 (figure  
28

6), and the system 101 will display a properties window 802 with a set of property values 803 which may be set.

A "minimum rank of returned hits" property 804 is a threshold value for including matched objects 106; matched objects 106 whose rank falls below this value are not displayed in the results window 602 and are not used in further processing. The rank of a matched object 106 is calculated by the CBR engine 110. In the example, this value is set to 80.

A "maximum clustered hits" property 805 is a maximum number of matched objects 106 which are included in a single cluster 402. Those matched objects 106 not included in clusters 402 are placed in a special cluster 402 labelled "Other". In the example, this value is set to 400.

A "clustering time" property 806 is the elapsed real time devoted to clustering. In the example, this value is set to 2500 milliseconds.

A "minimum number of clusters" property 807 is the lower bound for the number of clusters 402 generated. In the example, this value is set to 2 clusters.

A "maximum number of clusters" property 808 is the upper bound for the number of clusters 402 generated. In the example, this value is set to 8 clusters. The system 101

1 attempts to generate a number of clusters 402 between the minimum  
2 and maximum number selected.

3  
4 A "maximum topics" property 809 is the maximum number  
5 of topics displayed in the topics field 609 in the query window  
6 601. In the example, this value is set to 7 topics.

7  
8 A "maximum subtopics" property 810 is the maximum  
9 number of subtopics displayed in the subtopics window 610. In  
10 the example, this value is set to 250 subtopics.

11  
12 A "do/don't cluster" property 811 sets whether or not  
13 clustering is performed. In the example, this value is set to  
14 YES.

15  
16 A "do/don't generate query topics" property 812 sets  
17 whether or not topics and subtopics are generated in response to  
18 query terms 310. In the example, this value is set to YES.

19  
20 A "do/don't generate salient topics" property 813 sets  
21 whether or not topics and subtopics are generated in response to  
22 relevant words 211. In the example, this value is set to YES.

23  
24 A "boolean/vector query" property 814 sets whether the  
25 object file system 107 performs a boolean query or a vector query  
26 in response to the explorer 105. In the example, this value is  
27 set to vector queries. A boolean query would have boolean  
28 connectors (e.g., "AND", "OR") coupling the query terms 310, so

that the query 104 would not be as flexibly matched. Search using boolean queries is well known in the art.

## APPENDICES

Appendix A shows a table of parts of speech and a set of lexical rules for the English language, which may be used for the tag-and-segment-text process or the tag-and-segment-text process in a preferred embodiment.

Appendix B shows an output of a test run of an example filter when applied to a portion of an example multimedia encyclopedia used as a database, available as "Microsoft Encarta" from Microsoft Corporation of Redmond, Washington.

### Alternative Embodiments

While preferred embodiments are disclosed herein, many variations are possible which remain within the concept and scope of the invention, and these variations would become clear to one of ordinary skill in the art after perusal of the specification, drawings and claims herein.

## APPENDIX A

## LEX2.TXT

Number of original entries from LDOCE and WordNet:

2466 lines of the form: Ability: skill, faculty, aptitude  
 11624 total terms on the right (downward relationships)  
 Terms never have their parents as children (no loops)

Parts of speech represented:

A - Adjective	strong, vivid, real
ADV - Adverb	weakly, dimly, very
AUX - Auxiliary Verb	can, shall, will
AXN - AUX not	can't, won't doesn't
BE - be	is, are, be, was
BTH - PQT/Double Conj.	both
CLN - Colon	:
CMA - Comma	,
CON - Connective	and, or, but
CRD - Cardinal	three, 3.14, twenty-two
D - Determiner	the, a, that
DAT - Date &/or Time	Friday, 3:00, Christmas
DDC - D/Double Conj.	either, neither
DO - Do (aux)	do, did, does
ENS - End Of Sentence	., ?, !, ;
ETC - "And Others"	..., etc., et.al.
GEN - Genitive	his, her, their
HAV - Have (aux)	have, had, has, having
IJ - Interjection	oh, shucks, well
INF - Infinitive marker	to
N - Noun	frog, pride, year
NEG - Negation	not
ORD - Ordinal	first, 2nd, last
P - Preposition	by, around, with, from
PA - Open Paren	(, {, <
PD - Post Determiner	many, several, next,
PN - Proper Noun	Zippy, Brad Allen
PQL - Pre-Qualifier	quite, rather, such



## LEX2.TXT

POT - Pre-Quantifier nary, many, half, all  
 PRN - Pronoun him, she, we  
 PRT - Participial Verb running, thinking  
 QA - Quantifier/Article that, this  
 QL - Qualifier some, many, every,  
 QLP - Post-Qualifier enough, 'nuff, indeed  
 QN - Quantified Noun everybody, nothing  
 REN - Close Paren ), ], }, >  
 RP - Relative Pronoun that, which  
 SOS - Start of Sentence, or "  
 V - Verb (inf or past) eat, voted, surf  
 WHD - Wh-Determiner what, which  
 WHQ - Wh-Qualifier who, why  
 XT - Existential Term it, there

Total number of phrase recognition rules:

5 for the filter:

A|CRD|GEN|N|ORD, N, ^N  
 GEN, PRT  
 A|ADV|CRD|GEN|N|ORD, A|CRD|N|ORD, N, ^N  
 A|ADV|CRD|GEN|N|ORD, A|CRD|N|ORD, A|CRD|N|ORD, N, ^N  
 A|CRD|N|ORD, CON, A|CRD|N|ORD, N, ^N

Additional 10 for the Explorer (original 5 used as well):

N, P, N, ^N  
 N, RP, V, ^-(A|D|N|PN)  
 N, P, D|PD, N, ^N  
 N, RP, N|PN|V, N|PN|V  
 N, RP, AUX|AXN|COP|DO|HAV, A|PRT|V, ^-(A|CRD|D|N|ORD|PN)  
 A|ADV|CRD|GEN|N|ORD, A|CRD|N|ORD, A|CRD|N|ORD, N, ^N  
 N, P, D|PD, A|CRD|N|ORD, N, ^N  
 N, RP, V, A|CRD|D|N|ORD|PD, N  
 N, RP, A|CRD|D|N|ORD|PD, N, V

## LEX2.TXT

N, RP, AUX|AXN|COP|DO|HAV, P|PRT|V, N|PN

note: -X means not X or nothing at all (end of sentence)

Total number of automatically acquired lexicon entries:

For Encarta, including base LDOCE/Wordnet entries:

184904 unique words / base phrases

51623 parents involved in 445025 relationships

151850 children involved in 445025 relationships

Average number of terms per automatically acquired phrase:

445025 / 51623 = 8.6

445025 / 151850 = 2.9

Average number of children phrases from original LDOCE entries:

11624 / 2466 = 4.7

NOTE from Perry:

You asked how many things we got out of WordNet and LDOCE. The number that David responded was the number of taxonyms we extracted from those two sources (mostly WordNet). If you were asking the number of words we extracted, it was initially in the neighborhood of 85,000. The current number of tagged words in the lexicon is 25915.

There are some additional phrase lattice rules that David didn't mention, since they are currently stubbed out. They involve noun phrases where a prepositional phrase or relative clause attaches to the right of a noun:

Queen of England  
girl from Ipanema

## LEX2.TXT

man who hit Dave Adam  
car that didn't stop

The reason why we don't use them is because of the right attachment.  
Our current representation in the phrase lattice file is:

base-word, ext1, ext2, ... , extn

where ext1 through extn all attach to the LEFT of base-word. Bear in mind, of course, that unstubbing the code and fixing the reps of this file will add this form of phrase lattice entry, but it will also increase the size of the phrase lattice file (perhaps double it).

LDOCE is basically a dictionary of British English, so we found a lot of words we weren't familiar with, as well as a lot of double entries to account for American spellings (e.g. color and colour). The lexical categories we were able to extract out of LDOCE and WordNet were limited to nouns, verbs, adjectives, adverbs, conjunctions, determiners, predeterminers, prepositions, pronouns, and phrases. Since we don't use a phrasal lexicon, we threw the phrases away.

All other categories of words (including the different categories of verbs: do, be, have, participial) were hand tagged. This tagging was greatly aided by two books: DeRose's dissertation and the book by Kucera and Francis. The past tenses for all verbs were also done by hand, which was something of a waste as most of them (the regular ones) were eventually thrown away, once we implemented rules that tag based on word endings.

The following are the current set of rules used for determining noun phrases:

1. noun-phrase  $\rightarrow$  proper-noun (e.g. "Elvis")
2. noun-phrase  $\rightarrow$  pronoun (e.g. "he")
3. noun-phrase  $\rightarrow$  noun (e.g. "cars")
4. noun-phrase  $\rightarrow$  gerund (e.g. "running")
5. noun-phrase  $\rightarrow$  determiner noun-phrase (e.g. "The person")
6. noun-phrase  $\rightarrow$  quantifier noun-phrase (e.g. "Three people")
7. noun-phrase  $\rightarrow$  adjective noun-phrase (e.g. "fluffy clouds")
8. noun-phrase  $\rightarrow$  adverb noun-phrase (e.g. "maddeningly fluffy clouds")
9. noun-phrase  $\rightarrow$  noun noun-phrase (e.g. "printer ribbons")
10. noun-phrase  $\rightarrow$  noun-phrase relative-clause (e.g. "The car that hit me")
11. noun-phrase  $\rightarrow$  noun-phrase prepositional-phrase  
(e.g. "The person with the most toys")
12. noun-phrase  $\rightarrow$  noun-phrase that sentence  
(e.g. "The candidate that I will vote for")
13. noun-phrase  $\rightarrow$  noun-phrase [, noun-phrase]\* [,] and noun-phrase  
(e.g. "Larry, Moe and Curly")
14. noun-phrase  $\rightarrow$  noun-phrase [, noun-phrase]\* [,] or noun-phrase  
(e.g. "England, France, or Germany")
15. noun-phrase  $\rightarrow$  comparative noun-phrase than noun-phrase  
(e.g. "more tea than China")

The *Find Taxonomic Relations* process (process 2.2 in figure 4) uses ART-IM rules to capture patterns of words which indicate taxonomic relationships between the words. For example, it detects patterns like:

"... kangaroos, wallabies, and other marsupials ..."

From this particular phrase, one could reasonably extract the relations

IS\_A(kangaroo,marsupial) and  
IS\_A(wallaby,marsupial)

Other patterns which detect this type of relation extracted from [14] are :

1. NP such as (NP,)\* {(and | or) } NP
2. such NP as (NP,)\* {(and | or) } NP
3. NP (, NP)\* (,) and other NP
4. NP (, NP)\* (,) or other NP
5. NP (,) including (NP,)\* {(and | or) } NP
6. NP (,) especially (NP,)\* {(and | or) } NP

## APPENDIX B

Mar 16 17:39 1993 test.log Emacs buffer Page 1

Clustering file afl.txt

Non-empty clusters: 5

Clusters: 5

# Hits Vals Seed, Value:Count

```

0 1 0 NONE
1 2 0 Reuther, Walter Philip, Labor, labor:2, president:2, wage:2
2 2 0 Railroad Labor Organizations, Brotherhood, Union, united states:2,
3 7 0 Hillman, Sidney, Labor, labor:7, afl:7, union:4, american federati
4 2 0 Kirkland, Lane, Labor, director:2

```

Passes: 1029, best pass: 830, best score: 0.955, worst score: 0.170

Cluster 0, has 1 hits: ''

Football, Type, United States

Cluster 1, has 2 hits: 'labor:2, president:2, wage:2'

Meany, George, Labor

Reuther, Walter Philip, Labor

Cluster 2, has 2 hits: 'united states:2, union:2, management:2'

Railroad Labor Organizations, Brotherhood, Union

Teamsters Union, Full, International Brotherhood

Cluster 3, has 7 hits: 'labor:7, afl:7, union:4, american federation:4, cio:3, o

American Federation, Labor, Congress

Gomper, Samuel, Labor

Green, William, Labor

Hillman, Sidney, Labor

Knight, Labor, Union

Lewi, John L, Labor

Strike, Labor, Relation

Cluster 4, has 2 hits: 'director:2'

Kirkland, Lane, Labor

Rozelle, Pete, Full

Clustering file alcohol.txt

Non-empty clusters: 5

Clusters: 5

# Hits Vals Seed, Value:Count

```

0 15 0 (OTHER), blood:3, vitamin:2, tissue:2, poison:2, sugar metabolism:
1 22 0 Antifreeze, Chemical, Substance, alcohol:21, acid:7, ethyl:7, liqu
2 10 0 Vodka, Beverage, Known, alcohol:9, percent:5, beverage:5, use:3, l
3 6 0 Gasohol, Blend, Part, fuel:5, alcohol:2, methanol:2, combustion:2,
4 4 0 Marijuana, Mixture, Leave, drug:3, alcohol:3, syndrome:3, psychoac

```

Passes: 334, best pass: 158, best score: 0.307, worst score: 0.132

Cluster 0, has 15 hits: '(OTHER), blood:3, vitamin:2, tissue:2, poison:2, sugar

Birth Defects, Disorder, Structure

Cancer, Medicine, Growth

Corn, Maize, Cereal

Crop Farming, Cultivation, Plant

First Aid, Emergency, Measure

Fungi, Group, Organism

Liver, Organ, Vertebrate

Nutrition, Human, Science

Paint, Varnish, Liquid

Pennsylvania, Full, Commonwealth

Poison, Substance, Produce

Sugar, Term, Number

Mar 16 17:39 1993 test.log Emacs buffer Page 2

Thermometer, Instrument, Measure  
 Wine, Beverage, Juice  
 Wood, Substance, Trunk  
 Cluster 1, has 22 hits: 'alcohol:21, acid:7, ethyl:7, liquid:4, example:3, chemi  
 Acetaldehyde, Volatile, Liquid  
 Antifreeze, Chemical, Substance  
 Azeotropic Mixture, Solution, Ratio  
 Butyl Alcohol, Chemical, Formula  
 Cannizzaro, Stanislao, Italian  
 Disease, Medicine, Health  
 Ester, Chemistry, Compound  
 Ether, Chemistry, Ethyl  
 Fermentation, Chemical, Change  
 Formaldehyde, Compound, Carbon  
 Glycerin, Glycerol, C3h8o3  
 Gum, Substance, Plant  
 Iodine, Element, Symbol  
 Lipid, Group, Substance  
 Salicylic Acid, White, Solid  
 Solution, Chemistry, Mixture  
 Tannin, Acid, Name  
 Turpentine, Name, Semifluid  
 Vinegar, Condiment, Preservative  
 Wax, Name, Ester  
 Whiskey, Liquor, Mash  
 Zymology, Zymurgy, Biochemistry  
 Cluster 2, has 10 hits: 'alcohol:9, percent:5, beverage:5, use:3, liquor:3, dist  
 Beer, Term, Beverage  
 Cider, Sweet, Juice  
 Cosmetic, Term, Preparation  
 Distillation, Process, Liquid  
 Distilled Liquors, Beverage, Alcohol  
 Gin, Liquor, Grain  
 Liqueur, Beverage, Spirit  
 Police, Agency, Community  
 Prohibition, Ban, Manufacture  
 Vodka, Beverage, Known  
 Cluster 3, has 6 hits: 'fuel:5, alcohol:2, methanol:2, combustion:2, coal:2, eng  
 Alcohol, Arabic, Al-kuhul  
 Automobile, Greek, Auto  
 Combustion, Process, Oxidation  
 Energy Supply, World, Resource  
 Gasohol, Blend, Part  
 Rocket, Term, Propulsion  
 Cluster 4, has 4 hits: 'drug:3, alcohol:3, syndrome:3, psychoactive drugs:2, mar  
 Alcoholism, Illness, Ingestion  
 Drug Dependence, State, Compulsion  
 Marijuana, Mixture, Leave  
 Psychoactive Drugs, Chemical, Substance

```

Clustering file bulb.txt
Non-empty clusters: 5
Clusters: 5
# Hits Vals Seed, Value:Count
-----

```

Mar 16 17:39 1993 test.log Emacs buffer Page 3

```

0 9 0 (OTHER), plant:3, united states:2, seed:2, gardening:2, flower:2
1 10 0 Radiometer, Instrument, Intensity, bulb:7, light:4, tuber:3, stem:
2 3 0 Electric Lighting, Illumination, Mean, lamp:3, glass:2, neon:2, ar
3 5 0 Autumn Crocus, Name, Herb, bulb:5, liliaceae:4, herb:3, lily:3, pi
4 6 0 Hygrometer, Type, Instrument, temperature:4, atmosphere:3, point:3
Passes: 598, best pass: 333, best score: 0.491, worst score: 0.208
Cluster 0, has 9 hits: '(OTHER), plant:3, united states:2, seed:2, gardening:2,
Disease, Plant, Deviation
Gardening, Cultivation, Plant
Garlic, Name, Herb
Genetics, Study, Trait
Gopher, French, Gauffre
Horticulture, Latin, Hortu
Peanut Worm, Name, Small
Spice, Flavoring, Part
Technology, Term, Process
Cluster 1, has 10 hits: 'bulb:7, light:4, tuber:3, stem:3, rhizome:3, electron:2
Bulb, Mass, Leave
Edison, Township, Middlesex County
Edison, Thomas Alva, Inventor
Onion, Name, Herb
Photoelectric Cell, Phototube, Electron
Photography, Technique, Permanent
Radiometer, Instrument, Intensity
Rhizome, Stem, Organ
Tuber, Stem, Plant
Ray, Radiation, Wavelength
Cluster 2, has 3 hits: 'lamp:3, glass:2, neon:2, arc:2, bulb:2, argon:2, light:2
Argon, Element, Symbol
Electric Lighting, Illumination, Mean
Neon Lamp, Glass, Bulb
Cluster 3, has 5 hits: 'bulb:5, liliaceae:4, herb:3, lily:3, pistil:2, height:2,
Autumn Crocus, Name, Herb
Hyacinth, Plant, Genu
Soap Plant, Amole, Native
Star-of-bethlehem, Name, Herb
Tuberose, Herb, Polianth
Cluster 4, has 6 hits: 'temperature:4, atmosphere:3, point:3, humidity:2, bulb:2
Blood Pressure, Pressure, Blood
Humidity, Moisture, Content
Hygrometer, Type, Instrument
Meteorology, Study, Atmosphere
Thermometer, Instrument, Measure

```



Vapor, Physic, Term

Clustering file columbus.txt

Non-empty clusters: 7

Clusters: 7

# Hits Vals Seed, Value:Count

#	Hits	Vals	Seed, Value:Count
0	4	0	(OTHER), century:2
1	4	0	Pinzn, Name, Family, expedition:3, voyage:2, hispaniola:2, pinta:2
2	5	0	Puerto Rico, Commonwealth, Spanish Estado Libre Asociado, spanish:
3	2	0	Samana Cay, Island, Bahama, atlantic ocean:2, landfall:2, san salv
4	6	0	Mississippi, East South Central, U.S., state:5, river:3, city:3, a

Mar 16 17:39 1993 test.log Emacs buffer Page 4

5 5 0 Santiago, Dominican Republic, Name, cacao:3, city:3, caribbean:2,  
 6 4 0 South America, Continent, Asia, death valley:2, south:2, slavery:2  
 Passes: 614, best pass: 65, best score: 0.520, worst score: 0.189  
 Cluster 0, has 4 hits: '(OTHER), century:2'  
 American Literature, Literature, English  
 Coln, Geography, City  
 Europe, Continent, World  
 Knight, Columbu, Organization  
 Cluster 1, has 4 hits: 'expedition:3, voyage:2, hispaniola:2, pinta:2, ship:2'  
 Columbu, Christopher, Italian Cristoforo Colombo  
 Pinzn, Name, Family  
 Ship, Type, Construction  
 Velzquez, Diego, Soldier  
 Cluster 2, has 5 hits: 'spanish:4, island:3, spain:2, de:2, christopher columbus  
 Bobadilla, Francisco, De  
 Cuba, Island, West Indies  
 Dsirade, Island, West Indies  
 Ferdinand V, The Catholic, King  
 Puerto Rico, Commonwealth, Spanish Estado Libre Asociado  
 Cluster 3, has 2 hits: 'atlantic ocean:2, landfall:2, san salvador:2, island:2,  
 Samana Cay, Island, Bahama  
 Cluster 4, has 6 hits: 'state:5, river:3, city:3, american civil war:2, ohio:2,  
 Columbu, Georgia, City  
 Columbu, Mississippi, City  
 Columbu, Ohio, City  
 Georgia, State, South Atlantic  
 Mississippi, East South Central, U.S.  
 Ohio, East North Central, U.S.  
 Cluster 5, has 5 hits: 'cacao:3, city:3, caribbean:2, dominican:2, santiago:2, c  
 Columbu, Indiana, City  
 Santiago, Dominican Republic, Name  
 Santo Domingo, Trujillo, City  
 Spanish Town, City, Jamaica  
 Tobago, Republic, Commonwealth  
 Cluster 6, has 4 hits: 'death valley:2, south:2, slavery:2, brazil:2, continent:  
 Black, America, Immigration

North America, Continent, Canada  
 South America, Continent, Asia  
 United States, America, Republic

Clustering file dualism.txt

Non-empty clusters: 5

Clusters: 5

#	Hits	Vals	Seed, Value:Count
0	2	0	NONE
1	5	0	Dualism, Philosophy, Theory, mind:5, philosopher:5, philosophy:3,
2	3	0	Devil, Hebrew, Belief, evil:3, god:3, good:2, human:2, middle ages
3	3	0	Paulician, Church, History, dualism:3, sect:3, bogomils:2, old tes
4	2	0	Docetism, Christian, Heresy, doctrine:2, human:2

Passes: 1050, best pass: 312, best score: 1.003, worst score: 0.397

Cluster 0, has 2 hits: ''

Austria, German, sterreich

Zoroastrianism, Religion, Persia

Mar 16 17:39 1993 test.log Emacs buffer Page 5

Cluster 1, has 5 hits: 'mind:5, philosopher:5, philosophy:3, matter:3, universe:

Dualism, Philosophy, Theory  
 Metaphysics, Branch, Philosophy  
 Monism, Greek, Mono

Occasionalism, Term, System

Philosophy, Greek, Philosophia

Cluster 2, has 3 hits: 'evil:3, god:3, good:2, human:2, middle ages:2, middle ea

Albigens, Follower, Single

Devil, Hebrew, Belief

Evil, Wrong, Harm

Cluster 3, has 3 hits: 'dualism:3, sect:3, bogomils:2, old testament:2, century:

Basillide, Teacher, Alexandria

Bogomils, Member, Sect

Paulician, Church, History

Cluster 4, has 2 hits: 'doctrine:2, human:2'

Docetism, Christian, Heresy

Neoplatonism, Designation, Doctrine

Clustering file infant.txt

Non-empty clusters: 7

Clusters: 7

#	Hits	Vals	Seed, Value:Count
0	4	0	NONE
1	3	0	Gesell, Arnold Lucius, Psychologist, infant:3, development:2
2	2	0	Incubator, Apparatu, Chamber, growth:2
3	2	0	Pregnancy, Childbirth, Term, birth:2, pregnancy:2, infant:2, child
4	2	0	Hondura, Republic, Central America, country:2, 1980s:2
5	3	0	Baptism, Greek, Baptein, rite:2, baptism:2
6	2	0	Japan, Japanese Dai, Great, manchuria:2, government:2, party:2

Passes: 835, best pass: 7 best score: 0.795, worst score: 0.274

Cluster 0, has 4 hits:

Free Trade, Interchange, Frontier  
Human, Name, Individual  
Perception, Process, Stimulation  
Scotland, Division, Kingdom

Cluster 1, has 3 hits: 'infant:3, development:2'

Gesell, Arnold Lucius, Psychologist  
Infancy, Period, Birth  
Sudden Infant Death Syndrome, Sld, Death

Cluster 2, has 2 hits: 'growth:2'

Incubator, Apparatu, Chamber  
Population, Term, Human

Cluster 3, has 2 hits: 'birth:2, pregnancy:2, infant:2, childbirth:2, women:2'

Obstetrics, Branch, Medicine  
Pregnancy, Childbirth, Term

Cluster 4, has 2 hits: 'country:2, 1980s:2'

Hondura, Republic, Central America  
Sierra Leone, Nation, Africa

Cluster 5, has 3 hits: 'rite:2, baptism:2'

Baptism, Greek, Baptein  
Circumcision, Removal, Part  
Mennonite, Religious, Group

Cluster 6, has 2 hits: 'manchuria:2, government:2, party:2'

China, Chinese Zhonghua Renmin Gongheguo, People Republic

Mar 16 17:39 1993 test.log Emacs buffer Page 6

Japan, Japanese Dai, Great

Clustering file israel.txt

Non-empty clusters: 4

Clusters: 4

# Hits Vals Seed, Value:Count

#	Hits	Vals	Seed, Value:Count
0	22	0	(OTHER), government:6, war:4, century:3, french revolution:3, coun
1	66	0	Judah, Old Testament, Name, israel:64, judah:20, old testament:20,
2	39	0	Nasser, Gamal Abdel, Egyptian, israel:32, arab:26, israeli:20, pal
3	11	0	Song, Solomon, Book, book:10, old testament:9, israel:9, chap:5, b

Passes: 127, best pass: 117, best score: 0.213, worst score: 0.083

Cluster 0, has 22 hits: '(OTHER), government:6, war:4, century:3, french revolut

Achille Lauro, Italian, Cruise  
Anti-semitism, Social, Agitation  
Asia, Continent, Island  
Assyria, Ashur, Ashshur  
Bahai, Persian, Glory  
Buber, Martin, Religious  
Cabala, Hebrew, Tradition  
Crusade, Expedition, Undertaken  
Eschatology, Discourse, Last  
Espionage, Collection, Information  
Iran, Islamic Republic, Republic

Jewish Art, Architect  
 Jewish Music, Religious Music  
 Nationalism, History, Movement  
 Portuguese Literature, Literature, Portuguese  
 Refugee, Person, Country  
 Romania, Republic, Europe  
 Saudi Arabia, Monarchy, Southwest Asia  
 Union, Soviet Socialist Republics, Russian Soyuz Sovyetskikh Sotsialisticheskii  
 United Nations, Organization, Nation-state  
 United States, America, Republic  
 Woman Suffrage, Right, Women  
 Cluster 1, has 66 hits: 'israel:64, judah:20, old testament:20, king:18, bc:12,  
 Abner, Old Testament, Cousin  
 Ahab, King, Israel  
 Amaziah, Hebrew, King  
 Ammonite, People, Region  
 Amo, Book, Old Testament  
 Angel, Greek, Aggelo  
 Apostle, Greek, Apostolo  
 Ashqelon, Town, Palestine  
 Balaam, Old Testament, Prophet  
 Kokhba, Simon, Name  
 Bene Israel, Community, Jew  
 Ben-zvi, Itzhak, Second  
 Bethlehem, Jordan, Hebrew  
 Bible, Holy Bible, Book  
 Carmel, Mount, Mountain  
 Diaspora, Greek, Dispersion  
 David, King, Bc  
 Edom, Old Testament, Times  
 Elat, Eilat, City

Mar 16 17:39 1993 test.log Emacs buffer Page 7

Elia, Century, Bc  
 Elisha, Old Testament, See  
 Ephraim, Hebrew, Old Testament  
 Esdraelion, Plain, Jezreel  
 Ezekiel, Book, Old Testament  
 Falasha, Sect, Ethiopia  
 Galilee, Galil, Circle  
 Gideon, Hebrew, Hewan  
 Habima Theater, Former, Name  
 Hebron, City, Israeli-occupied Jordan  
 Herzog, Chaim, President  
 High Priest, Hierarchy, Head  
 Holon, City, Israel  
 Israel, Kingdom, Hebrew  
 Jacob, Old Testament, Patriarch  
 Joash, Name, King  
 Jehoshaphat, Hebrew, Jehovah  
 Jehu, Hebrew, Jehovah  
 Jeremiah, Book, Old Testament

Jeroboam I, Old Testament, See  
 Jeroboam II, King, Israel  
 Jew, Usage, Hebrews  
 Jezebel, Tyrian, Princess  
 Jonathan, Old Testament Books, Samuel  
 Judah, Old Testament, Name  
 Judaism, Culture, Jew  
 Justification, Theology, Way  
 King, Book, Old Testament  
 Lost Tribes, History, Tribe  
 Manasseh, Son, Old Testament  
 Meir, Golda, Israeli  
 Michael, Hebrew, God  
 Moab, Country, Hill  
 National Jewish Welfare Board, National, Agency  
 Negeb, Region, Middle East  
 Philistine, Inhabitant, Region  
 Putnam, Israel, Soldier  
 Ramat Gan, City, Central  
 Rehoboam, King, Judah  
 Samuel, Book, Old Testament  
 Saul, King, Israel  
 Sharon, Plain, Israel  
 Shema, Hebrew, Word  
 Solomon, King, Israel  
 Tiberia, Lake, Sea  
 Weizmann, Chaim, Long-time  
 Zangwill, Israel, English  
 Cluster 2, has 39 hits: 'israel:32, arab:26, israeli:20, palestine:11, egypt:11,  
 Husein, King, Jordan  
 Acre, Akko, Seaport  
 Agnon, Shmuel Yosef, Israeli  
 Amman, Rabbah Ammon, Philadelphia  
 Arab League, Name, League  
 Arafat, Yasir, Palestinian  
 Aren, Moshe, Israeli  
 Menachem, Israeli, Prime

Mar 16 17:39 1993 test.log Emacs buffer Page 8

Ben-gurion, David, Israeli  
 Damascu, Arabic Dimashq, Ash-sham  
 Dayan, Moshe, Israeli  
 Egypt, Arab Republic, United Arab Republic  
 Gaza, Arabic Ghazze, City  
 Golan Heights, Region, Syria  
 Haifa, City, Seaport  
 Hebrew Literature, Literature, Jew  
 Iraq, Irak, Republic  
 Israel, Republic, Middle East  
 Jerusalem, Arabic, Al-qud  
 Jordan, River, Middle East  
 Jordan, Hashemite Kingdom, Arabic

Kibbutz, Village, Far  
 Lebanon, Arabic Lubnan, Republic  
 Libya, Full, Socialist People Libyan Arab Jamahiriyah  
 Middle East, Region, Geography  
 Nasser, Gamal Abdel, Egyptian  
 Palestine, Region, Extent  
 Palestine Liberation Organization, Plo, Body  
 Sadat, Egyptian, Military  
 Six-day War, Conflict, June  
 Suez Canal, Waterway, Running  
 Syria, Arabic Suriyah, Al-arabiyah  
 Tel Aviv-jaffa, Tel Aviv-yafa, City  
 Terrorism, International, Use  
 Tunisia, Republic, Africa  
 West Bank, Area, West  
 Yom Kippur War, Conflict, Israel  
 Zionism, Movement, People  
 Zionist Organization, America, Zoa  
 Cluster 3, has 11 hits: 'book:10, old testament:9, israel:9, chap:5, bc:5, proph  
 Dead Sea Scrolls, Collection, Hebrew  
 Hosea, Book, Old Testament  
 Isaiah, Book, Old Testament  
 Joshua, Book, Old Testament  
 Judge, Book, Old Testament  
 Micah, Book, Old Testament  
 Number, Book, Old Testament  
 Obadiah, Book, Old Testament  
 Song, Solomon, Book  
 Wisdom, Solomon, Book  
 Zechariah, Book, Old Testament

Clustering file marx.txt

Non-empty clusters: 6

Clusters: 6

# Hits Vals Seed, Value:Count

#	Hits	Vals	Seed, Value:Count
0	2	0	(OTHER), german:2, germany:2, east:2, baltic sea:2
1	3	0	Hegel, G, W, philosopher:3, philosophy:2
2	4	0	Bolshevism, Doctrine, Theory, communist:4, lenin:4, revolution:3,
3	4	0	Marx Brothers, 20th-Century, Comedian, marx:4, socialism:2, engels
4	4	0	Communist Manifesto, German Manifest, Partei, capitalist:3, class:
5	6	0	Ideology, System, Concept, social:3, marx:3, labor:2, world war ii

Mar 16 17:39 1993 test.log Emacs buffer Page 9

Passes: 722, best pass: 675, best score: 0.663, worst score: 0.248

Cluster 0, has 2 hits: '(OTHER), german:2, germany:2, east:2, baltic sea:2'

Germany, Country, Europe

Germany, German Democratic Republic, Gdr

Cluster 1, has 3 hits: 'philosopher:3, philosophy:2'

Hegel, G, W

Philosophy, Greek, Philosophia

Political Theory, Sub .v ion, Science  
 Cluster 2, has 4 hits: 'communist:4, lenin:4, revolution:3, communism:2, governm  
 Bolshevism, Doctrine, Theory  
 Communism, Concept, System  
 International, Name, Socialist  
 Socialism, Doctrine, Movement  
 Cluster 3, has 4 hits: 'marx:4, socialism:2, engels:2'  
 Bernstein, Eduard, German Social Democratic  
 Economics, Science, Production  
 Engels, Friedrich, German  
 Marx Brothers, 20th-century, Comedian  
 Cluster 4, has 4 hits: 'capitalist:3, class:3, capitalism:2, communist:2, bourge  
 Bourgeoisie, Resident, European  
 Capitalism, System, Individual  
 Communist Manifesto, German Manifest, Partei  
 Marx, Karl, German  
 Cluster 5, has 6 hits: 'social:3, marx:3, labor:2, world war ii:2, german:2, cen  
 Ideology, System, Concept  
 Karl-marx-stadt, Former, Name  
 Kautsky, Karl Johann, German Marxist  
 Lassalle, Ferdinand, German  
 Sociology, Science, Deal  
 Wage, Theory, Labor

Clustering file muslim.txt

Non-empty clusters: 4

Clusters: 4

# Hits Vals Seed, Value:Count

#	Hits	Vals	Seed	Value:Count
0	41	0	(OTHER), arab:7, bc:5, ibn:4, indian:4, india:4, islam:4	
1	20	0	Philippine, Republic, Pacific Ocean, 1980s:17, country:8, governme	
2	40	0	Kashgar, Kashi, Kaxgar, muslim:38, india:8, muhammad:7, jerusalem:	
3	11	0	Mathematics, Study, Relationship, century:11, art:3, france:3, arc	

Passes: 146, best pass: 47, best score: 0.210, worst score: 0.124  
 Cluster 0, has 41 hits: '(OTHER), arab:7, bc:5, ibn:4, indian:4, india:4, islam:  
 Alfonso VIII, King, Castile  
 Arabia, Desert, Peninsula  
 Arabic Literature, Literature, People  
 Archaeology, Greek, Archaio  
 Averros, Arabic, Abu  
 Black Muslims, Religious, Organization  
 Borneo, Island, World  
 Chess, Game, Skill  
 Christianity, World, Religion  
 Chronology, Science, Division  
 Concubinage, Term, World  
 Costume, Clothing, People  
 Demon, Usage, Spirit

Mar 16 17:39 1993 test.log Emacs buffer Page 10

Egypt, Arab Republic, United Arab Republic

Gandhi, Mohandas Karar . . 1, Mahatma Gandhi  
 Ghana, Kingdom, West African  
 Hegira, Hejira, Arabic  
 Iraq, Irak, Republic  
 Jacobite Church, Christian, Group  
 Java, Island, Malay Archipelago  
 Jew, Usage, Hebrews  
 Jordan, Hashemite Kingdom, Arabic  
 Judaism, Culture, Jew  
 Karbala, City, Iraq  
 Mahdi, Arabic, Mahdiy  
 Medina, Medinat-en-nabi, City  
 Middle East, Region, Geography  
 Nehru, Indian, Nationalist  
 Orthodox Church, Major, Branch  
 Philosophy, Greek, Philosophia  
 Pottery, Clay, Firing  
 Punjab, Region, River  
 Saudi Arabia, Monarchy, Southwest Asia  
 Shiite, Arabic, Partisan  
 Sikhs, Follower, Religion  
 Sudan, Republic, Africa  
 Trigonometry, Branch, Mathematics  
 Tobago, Republic, Commonwealth  
 Tunisia, Republic, Africa  
 Turkey, Republic, Turkish Trkiye Cumhuriyeti  
 Vijayanagar, Kingdom, India  
 Cluster 1, has 20 hits: '1980s:17, country:8, government:7, spanish:5, arab:4, s  
 Afghanistan, Persian Afghnistn, Republic  
 Bangladesh, Full, People Republic  
 Berber, Name, Language  
 Cameroon, Republic, Africa  
 Chad, Republic, Central  
 Ethiopia, Abyssinia, Republic  
 Gambia, Republic, Commonwealth  
 Gibraltar, Dependency, Promontory  
 Indonesia, Republic, Island  
 Iran, Islamic Republic, Republic  
 Israel, Republic, Middle East  
 Kenya, Republic, Africa  
 Libya, Full, Socialist People Libyan Arab Jamahiriyah  
 Morocco, Arabic, Al-mamlakah  
 Nigeria, Federal Republic, Republic  
 Pakistan, Islamic Republic, Republic  
 Philippine, Republic, Pacific Ocean  
 Republic, Europe, Portion  
 Spain, Spanish Espaa, Monarchy  
 Syria, Arabic Suriyah, Al-arabiyah  
 Cluster 2, has 40 hits: 'muslim:38, india:8, muhammad:7, jerusalem:5, delhi:4, p  
 Fakhruddin Ali, Fifth, President  
 Algeria, French Algrie, Popular Republic  
 Allah, Name, Supreme Being  
 Almeida, Francisco, De  
 Almoravid, Berber, Dynasty  
 Asia, Continent, Island



Mar 16 17:39 1993 test.log Emacs buffer Page 11

Babism, Religion, Offshoot  
 Balewa, Sir Abubakar Tafawa, Minister  
 Region, Part, Subcontinent  
 Caliphate, Office, Realm  
 Crusade, Expedition, Undertaken  
 Delhi, Old Delhi, City  
 Delhi Sultanate, Muslim, State  
 Dervish, Turkish, Darvsh  
 Fakir, Arabic, Faqir  
 Farabi, Tarkhan, Al-farabi  
 Gansu, Kansu, Province  
 Ghazali, Name, Abu Hamid Muhammad  
 India, Republic, Hindi Bharat  
 Sir Muhammad, Pakistani, Philosopher  
 Islam, World, Religion  
 Islamic Music, Vocal, Art  
 Jammu, Kashmir, Known  
 Jerusalem, Arabic, Al-qud  
 Jinnah, Muhammad Ali, Leader  
 Kashgar, Kashi, Kaxgar  
 Kharijite, Arabic, Kharawrij  
 Lebanon, Arabic Lubnan, Republic  
 Malaysia, Monarchy, Commonwealth  
 Malcolm X, Leader, Omaha  
 Mufti, Title, Lawyer  
 Palestine, Region, Extent  
 Pilgrim, Place, Intent  
 Relic, Usage, Body  
 Roger I, Norman, Conqueror  
 Saladin, Leader, Jerusalem  
 Shivaji Bhonsle, Founder, India Maratha State  
 Tughluq, Muhammad, Sultan  
 Tuni, Tune, City  
 Umar, Al-hajj, West African  
 Cluster 3, has 11 hits: 'century:11, art:3, france:3, architecture:2, sculpture:  
 Africa, Continent, Island  
 Europe, Continent, World  
 France, French Rpublique Franaise, Republic  
 Gypsy, People, Heritage  
 History, Historiography, Sense  
 Indian Art, Architecture, Art  
 Indian Literature, Literature, Language  
 Islamic Art, Architecture, Art  
 Library, Repository, Form  
 Mathematics, Study, Relationship  
 Portraiture, Representation, Art

Clustering file pope.txt

Non-empty clusters: 3

Clusters: 3

! Hits Vals Seed, Value:Count

	0	50	0 (OTHER), church:12, henry:8, king:7, english:6, roman:6, governmen
1	138	0	Benedict Xiv, Pope, Moderation, pope:138, church:28, rome:26, coun

2 12 0 Angelico. r Italian, florence:10, megalic, florentine:4, domin

Mar 16 17:39 1993 test.log Emacs buffer Page 12

Passes: 86, best pass: 34, best score: 0.149, worst score: 0.082  
 Cluster 0, has 50 hits: '(OTHER), church:12, henry:8, king:7, english:6, roman:6  
 Aquina, Saint Thomas, Angelic Doctor  
 Borgia, Cesare, Italian  
 Bruno, Saint, Carthusian  
 Bulgaria, Full, People Republic  
 Canon Law, Greek, Kanon  
 Carpin, Giovanni, De  
 Carroll, John, American Roman Catholic  
 Christianity, World, Religion  
 Church, England, Anglican Church  
 Civil War, Conflict, United States  
 Conrad III, King, Germany  
 Corsica, French Corse, Island  
 Counter Reformation, Movement, Roman Catholic  
 Couplet, Poetry, Term  
 Cranmer, Thoma, Archbishop  
 Cyril, Methodiu, Saint  
 Demarcation, Line, Boundary  
 Duns Scotus, John, Theologian  
 Easter, Festival, Resurrection  
 England, Latin Anglia, Portion  
 English Literature, Literature, England  
 Erigena, John Scotus, Scholar  
 Este, Italian, Family  
 Europe, Continent, World  
 Felix V, Last, Antipope  
 Ferdinand I, Naple, King  
 Feuillant, French, Organizations-one  
 Finland, Finnish Suomi, Republic  
 Fisher, Saint John, English Christian  
 France, French Rpublique Franaise, Republic  
 Gardiner, Stephen, English  
 Germany, Country, Europe  
 Henry VIII, King, England  
 Henry IV, France, Bourbon  
 Holy Roman Empire, Entity, Europe  
 Hungary, Hungarian Magyarorszg, Republic  
 Ireland, Geography, Island  
 Italian Italia, Republic, Europe  
 Knight, Saint John, Jerusalem  
 Lincoln, Abraham, President  
 Loyola, Saint Ignatius, Spanish Inigo  
 Lutheranism, Protestant, Denomination  
 Mary, Virgin Mary, Mother  
 Mendelssohn, Mos, German  
 Middle Ages, Period, European  
 Modernism, Theology, Philosophy  
 Neri, Saint Philip, Italian

Orthodox Church, Major, Inch  
 Poland, Republic, Polska Rzeczpospolita  
 Pole, Reginald, English Roman Catholic  
 Cluster 1, has 138 hits: 'pope:138, church:28, rome:26, council:23, papacy:23, h  
 Adrian I, Pope, Power  
 Adrian IV, Pope, Englishman  
 Adrian VI, Pope, Dutchman

Mar 16 17:39 1993 test.log Emacs-buffer Page 13

Alexander III, Pope, Authority  
 Alexander VI, Pope, Worldliness  
 Algardi, Alessandro, Italian  
 Antonelli, Giacomo, Italian  
 Arnold, Brescia, 1100-c  
 Augustinian, Order, Roman Catholic  
 Bacon, Roger, English Scholastic  
 Basel, Council, Middle Ages  
 Bembo, Pietro, Italian  
 Benedict VIII, Pope, Reformer  
 Benedict IX, Pope, 1032-44  
 Benedict XIII, Antipope, Avignon  
 Benedict XIV, Pope, Moderation  
 Benedict XV, Pope, Church  
 Bernard, Clairvaux, Saint  
 Bonaventure, Saint, Theologian  
 Boniface, Saint, English Benedictine  
 Boniface VIII, Pope, Power  
 Boniface IX, Pope, Papal States  
 Bossuet, Jacques Benigne, French Roman Catholic  
 Bull, Letter, Document  
 Bull Run, Battle, Manassas  
 Callistus, Calixtus I, Saint  
 Callistus II, Calixtus II, Pope  
 Callistus III, Calixtus III, Pope  
 Canonization, Roman Catholic, Church  
 Canossa, Village, Reggio  
 Cardinal, Title, Latin  
 Catherine, Aragon, Queen  
 Catherine, Siena, Saint  
 Cedar Mountain, Battle, Military  
 Celestine V, Saint, Pope  
 Celestine III, Pope, Born Giacinto Bobo  
 Censorship, Supervision, Control  
 Chalcedon, Council, Emperor  
 Charlemagne, Latin Carolus Magnus, Charles  
 Charles V, Holy Roman Empire, Holy Roman  
 Church, State, Relationship  
 Clement V, Pope, Avignon  
 Clement VI, Pope, Church  
 Clement VII, Pope, Pontificate  
 Clement VII, Antipope, Great Schism  
 Clement VIII, Last, Pope

Clement XIV, Pope, Jesui  
 Conciliar Theory, Doctrine, Superiority  
 Conclave, Latin, Cum  
 Constance, Council, City  
 Coptic Church, Christian, Church  
 Council, Assembly, Doctrine  
 Crusade, Expedition, Undertaken  
 Damasus I, Saint, Pope  
 Damian, Saint Peter, Doctor  
 Doctor, Church, Christian  
 Dllinger, Johann Joseph Ignaz, Von  
 Ecumenical Movement, Movement, Cooperation  
 Edmund, Abingdon, Saint

Mar 16 17:39 1993 test.log Emacs buffer Page 14

Elector, German Imperial, German Kurfrsten  
 Eugene Iii, Pope, Cistercian  
 Eugene Iv, Pope, Dispute  
 Formosu, Pope, Trial  
 Franciscan, Order, Friars Minor  
 Frederick I, Holy Roman Empire, Frederick Barbarossa  
 Frederick Ii, Holy Roman Empire, Holy Roman  
 Gallicanism, History, Combination  
 Gregory I, Saint, Pope  
 Gregory Ii, Saint, Pope  
 Gregory Vii, Saint, Pope  
 Gregory Ix, Pope, Inquisition  
 Gregory Xi, Pope, Return  
 Guiscard, Robert, Norman  
 Henry Ii, Holy Roman Empire, Henry The Saint  
 Henry Iv, Holy Roman Empire, Holy Roman  
 Henry V, Holy Roman Empire, German  
 Hippolytu, Rome, Saint  
 Honorius I. Pope, Heretic  
 Infallibility, Theology, Doctrine  
 Innocent Iii, Pope, Pop  
 Innocent Iv, Pope, Dominion  
 Innocent Xi, Pope, King Louis Xiv  
 Inquisition, Institution, Papacy  
 Interdict, Roman Catholic, Church  
 Investiture Controversy, Dispute, Church  
 Jesuit, Society, Jesu  
 Joan, Pope, Female  
 John Ii, Pope, Born Mercurius  
 John Viii, Pope, Ablest  
 John Xii, Pope, Boy Pope  
 John Xxi, Pope, Pontiff  
 John Xxii, Pope, Second  
 John Xxiii, Antipope, Born Baldassare Cossa  
 John Xxiii, Pope, Era  
 John, John Lackland, King  
 John Paul I, Pope, Born Albino Luciani

John Paul II, Pope, N . . . lian  
 Jubilee, Jew, Sabbatical  
 Julius II, Pope, Reign  
 Kulturkampf, German, Culture  
 Langton, Stephen, English  
 Lateran Councils, Council, Roman Catholic  
 Lateran Treaty, Designation, Agreement  
 Leo III, Saint, Pope  
 Leo IX, Saint, Pope  
 Leo X, Pope, Renaissance  
 Leo XIII, Pope, Modern  
 Louis IV, German, Ludwig IV  
 Lyon, Council, Church  
 Martin I, Saint, Pope  
 Martin IV, Pope, Born Simon  
 Martin V, Pope, Election  
 Molino, De, Spanish Roman Catholic  
 Nicholas III, Pope, Papal States  
 Nichola, Cusa, German

Mar 16 17:39 1993 test.log Emacs buffer Page 15

Occam, William, 1285-1349  
 Otto III, Holy Roman, Emperor  
 Otto IV, Otto, Brunswick  
 Papacy, Office, Pope  
 Papal States, Church, Pontifical States  
 Paschal II, Pope, Reign  
 Paul V, Pope, Born Camillo Borghese  
 Paul VI, Pope, Second Vatican Council  
 Pepin, Short, Mayor  
 Peter Pence, Offering, Pope  
 Philip IV, France, The Fair  
 Photiu, 820-91, Patriarch  
 Pico Della Mirandola, Giovanni, Conte  
 Pius II, Pope, Writer  
 Pius IV, Pope, Conclusion  
 Pius V, Saint, Pope  
 Pius VI, Pope, Reign  
 Pius VII, Pope, Napoleon  
 Pius IX, Pope, Pontificate  
 Pius X, Saint, Pope  
 Pius XI, Pope, Path  
 Pius XII, Pope, World War II  
 Pope, Latin, Papa  
 Cluster 2, has 12 hits: 'florence:10, medici:5, florentine:4, dominican:3, church  
 Alberti, Leon Battista, Italian  
 Albertus Magnus, Saint, Albert  
 Angelico, Fra, Italian  
 Cellini, Benvenuto, Florentine  
 Dante Alighieri, Italian, Poet  
 Dominican, Friars Preachers, Member  
 Ferrara-florence, Council, Basel-ferrara-florence

Florence, Italian Fire.2 Florentia  
 Guicciardini, Francesco, Italian  
 Leonardo, Da, Vinci  
 Medici, Lorenzo, De  
 Michelangelo, Creator, History

Clustering file sound.txt

Non-empty clusters: 5

Clusters: 5

# Hits Vals Seed, Value:Count

0	68	0 (OTHER), music:10, american civil war:6, state:6, bass:5, century:
1	57	0 Mach Number, Aerodynamics, Mechanic, sound:51, instrument:8, pitch
2	8	0 Letter, Vowel, English, sound:6, long:3, letter:3, sign:2, atlanti
3	19	0 Linguistics, Study, Language, language:14, english:9, speech:6, so
4	11	0 Vowel, English, Alphabet, sound:11, alphabet:9, letter:9, hierogly

Passes: 103, best pass: 74, best score: 0.173, worst score: 0.072

Cluster 0, has 68 hits: '(OTHER), music:10, american civil war:6, state:6, bass:

Amati, Family, Italian

American Indian Languages, Language, People

American Indians, People, America

Audiovisual Education, Planning, Preparation

Band, Ensemble, Brass

Transaction, Service, Consumer

Mar 16 17:39 1993 test.log Emacs buffer Page 16

Bird, Name, Member

Bremerton, City, Kitsap County

British Columbia, Province, Canada

Bronx, Borough, New York City

Building Construction, Procedure, Erection

Circulatory System, Anatomy, Physiology

Communication, Method, Receiving

Connecticut, New England, United States

Copyright, Body, Right

Currency, Economics, Term

Deep-sea Exploration, Investigation, Chemical

Bass, Member, Violin

Drama, Dramatic Arts, Form

Edison, Thomas Alva, Inventor

Encyclopedia, Encyclopaedia, Greek

Firework, Device, Material

Floor, Floor Coverings, Ceiling

Folk Dance, Dance, Member

Folk Music, Music, Performance

Frequency, Term, Science

Golden Globe Awards, Motion, Picture

Harmony, Music, Combination

Harpsichord, Italian, Cembalo

Insect, Name, Animal

Jazz, Type, Music

Jet Propulsion, Thrust, parting  
 Mississippi, East South Central, U.S.  
 Motion Picture Arts, Science, Academy  
 Music, Vocal, Part  
 Music, Western, Europe  
 Musical Form, Arrangement, Element  
 Mystic, Village, Stonington  
 Navigation, Science, Position  
 Haven, City, New Haven County  
 North Carolina, South Atlantic, U.S.  
 Ocean, Oceanography, Body  
 Orchestra, Ensemble, Instrument  
 Orchestration, Art, Musical  
 Philosophy, Greek, Philosophia  
 Pianoforte, Keyboard, Musical  
 Social Dance, Term, Dance  
 Radio, System, Communication  
 Rhode Island, Full, State  
 Scale, Music, Italian  
 Scott, Robert Falcon, Officer  
 Seattle, City, Seat  
 Seward Peninsula, Peninsula, Alaska  
 Snake, Reptile, Name  
 Sonata, Italian, Sonare  
 Tacoma, City, Seat  
 Telephone, Communication, Instrument  
 Television, Tv, Transmission  
 Theater Production, Mean, Form  
 United States, America, Republic  
 Valdez, City, Alaska  
 Video Recording, Process, Recording

Mar 16 17:39 1993 test.log Emacs buffer Page 17

Viol, Instrument, Century  
 Washington, State, U.S.  
 Wave Motion, Physic, Mechanism  
 Whale, Mammal, Order  
 Yachting, Operation, Boat  
 Zither, Instrument, String  
 Cluster 1, has 57 hits: sound:51, instrument:8, pitch:7, string:5, recording:5,  
 Acoustics, Greek, Akouein  
 Aerodynamics, Branch, Mechanic  
 Airplane, Craft, Action  
 Albemarle Sound, Inlet, Atlantic Ocean  
 Bell, Instrument, Percussion  
 Chaplin, Charlie, Name  
 Clair, Ren, Name  
 Digital Audio Tape, Dat, Tape  
 De Forest, Lee, Inventor  
 Doppler Effect, Physic, Variation  
 Ear, Organ, Hearing  
 Edmond, City, Snohomish County

Electronic Music, Music, Knowledge  
 Exxon Valdez, Oil, Tanker  
 Falkland Islands, Islas Malvinas, Island  
 Fluid Mechanics, Science, Action  
 Grunt, Name, Fish  
 Guitar, Instrument, Lute  
 Harmonic, Vibration, Primary  
 Harp, Instrument, Run  
 Hearing, Main, Sense  
 Hearing Aid, Device, Sound  
 Mach Number, Aerodynamics, Mechanic  
 Microphone, Device, Energy  
 Midi, Acronym, Musical Instrument Digital Interface  
 Motion Picture, Sequence, Photograph  
 Motion Pictures, History, Development  
 Music, Movement, Sound  
 Musical Instruments, Tool, Scope  
 Noise, Physic, Signal  
 Oboe, Wind, Instrument  
 Organ, Instrument, Air  
 Petroleum, Oil, Bituminous  
 Phonograph, Known, Player  
 Physic, Science, Constituent  
 Prince William Sound, Inlet, Gulf  
 Propeller, Device, Force  
 Puget Sound, Arm, Pacific Ocean  
 Radiometer, Instrument, Intensity  
 Reflection, Physic, Phenomenon  
 Singing, Use, Voice  
 Sonar, Acronym, Sound Navigation And Ranging  
 Sound, Phenomenon, Sense  
 Determination, Depth, Body  
 Sound Recording, Reproduction, Conversion  
 Supersonics, Branch, Physic  
 Synthesizer, Computer, Peripheral  
 Tone, Music, Sound  
 Transformer, Device, Coil

Mar 16 17:39 1993 test.log Emacs buffer Page 18

Tyndall, John, Physicist  
 Ultrasonics, Branch, Physic  
 Ventriloquism, Art, Sound  
 Violin, Instrument, Member  
 Viscount Melville Sound, Arm, Arctic Ocean  
 Voiceprint Identification, Method, Person  
 Warner Brothers, Motion, Picture  
 Xylophone, Greek, Xylon  
 Cluster 2, has 8 hits: 'sound:6, long:3, letter:3, sign:2, atlantic ocean:2, mi:  
 Animal Behavior The, Behavior, Animal  
 C, English, Romance-language  
 Diacritic Mark, Sign, Mark  
 Island Sound, Body, Salt



Letter, Vowel, English  
 Pamlico Sound, Inlet, Atlantic Ocean  
 Rhyme, Likeness, Sound  
 W, Letter, English  
 Cluster 3, has 19 hits: 'language:14, english:9, speech:6, sound:6, word:5, spok  
 American English, English, Spoken  
 Celtic Languages, Indo-european, Family  
 Chinese Language, Language, Chinese  
 Cuneiform, Latin, Cuneu  
 Deafness, Inability, Definition  
 English Language, Medium, Communication  
 English Literature, Literature, England  
 Etymology, Branch, Linguistics  
 Grammar, Branch, Linguistics  
 Greek Language, Language, People  
 Hieroglyph, Character, System  
 Japanese Language, Language, Spoken  
 Language, Communication, Being  
 Linguistics, Study, Language  
 Phonetics, Branch, Linguistics  
 Poetry, Form, Expression  
 Semantics, Greek, Semantiko  
 Versification, Art, Verse  
 Writing, Method, Intercommunication  
 Cluster 4, has 11 hits: 'sound:11, alphabet:9, letter:9, hieroglyph:8, english:7  
 Vowel, English, Alphabet  
 Alphabet, Alpha, Beta  
 F, Letter, Consonant  
 K, Letter, English  
 L, Letter, English  
 M, Letter, English  
 Q, Letter, English  
 R, Letter, English  
 U, 21st, Letter  
 X, Letter, English  
 Y, Letter, English

Clustering file strike.txt

Non-empty clusters: 4

Clusters: 4

# Hits Vals Seed, Value:Count

-----

Mar 16 17:39 1993 test.log Emacs buffer Page 19

0 6 0 (OTHER), electron:2, beam:2, tube:2, television:2  
 1 11 0 Gary, City, Lake County, strike:10, united states:3, president:2,  
 2 10 0 National Labor Relations Act, Nlra, Law, labor:9, strike:8, union:  
 3 15 0 Poland, Republic, Polska Rzeczpospolita, government:11, 1980s:8, w  
 Passes: 453, best pass: 208, best score: 0.445, worst score: 0.154  
 Cluster 0, has 6 hits: '(OTHER), electron:2, beam:2, tube:2, television:2'  
 Baseball, Game, Skill

Cathode-ray Tube, Elm, Tube  
 Napoleon I, Emperor, French  
 Russia, History, Empire  
 Television, Tv, Transmission  
 Warfare, Use, Force  
 Cluster 1, has 11 hits: 'strike:10, united states:3, president:2, injunction:2,  
 Chartism, Reform, Movement  
 Coolidge, John, Calvin  
 Defense Systems, Defense, Country  
 Deb, Eugene Victor, American Socialist  
 Dollfuss, Engelbert, Chancellor  
 Fault, Geology, Line  
 Gary, City, Lake County  
 Homestead Strike, Labor, Strike  
 Pullman Strike, See, Deb  
 Sound, Phenomenon, Sense  
 Ueberroth, Peter Victor, Sport  
 Cluster 2, has 10 hits: 'labor:9, strike:8, union:7, labor-management relations  
 Cleveland, Grover, 22d  
 Industrial Workers, World, Former  
 International Ladies, Garment Workers, Union  
 Knight, Labor, Union  
 Labor Relations, Transaction, Determination  
 Lockout, Labor, Relation  
 National Labor Relations Act, Nlra, Law  
 Labor, Relation, Practice  
 Strike, Labor, Relation  
 Trade Unions, United States, Labor  
 Cluster 3, has 15 hits: 'government:11, 1980s:8, war:6, country:4, soviet:3, par  
 Colombia, Republic, South America  
 France, French Rpublique Franaise, Republic  
 Ghana, Country, Africa  
 Britain, United Kingdom, Great Britain  
 Illinois, East North Central, U.S.  
 Italian Italia, Republic, Europe  
 Japan, Japanese Dai, Great  
 Northern Ireland, Part, United Kingdom  
 Poland, Republic, Polska Rzeczpospolita  
 Russian Revolution, Event, Russia  
 Spain, Spanish Espaa, Monarchy  
 Sweden, Konungariket Sverige, Kingdom  
 Union, Soviet Socialist Republics, Russian Soyuz Sovyetskikh Sotsialisticheskii  
 United States, America, Republic  
 World War II, Military, Conflict

Clustering file utah.txt  
 Non-empty clusters: 5  
 Clusters: 5

Mar 16 17:39 1993 test.log Emacs buffer Page 20

# Hits Vals Seed, Value:Count

```

0 2 0 (OTHER), state:2
1 3 0 Utah, University, Institution, utah:3
2 9 0 City, Davis County, Utah, city:8, utah:8, mormon:5, state:4, name:
3 3 0 Mormonism, World, Religion, mormonism:3, polygamy:3, smith:3, morm
4 7 0 Green, River, Utah, utah:6, colorado:5, mi:4, km:4, river:2, yampa
Passes: 764, best pass: 515, best score: 0.652, worst score: 0.147
Cluster 0, has 2 hits: '(OTHER), state:2'
    United States, America, Republic
    State, U.S., North
Cluster 1, has 3 hits: 'utah:3'
    Bushnell, Nolan Kay, Founder-chairman
    Orem, City, Utah County
    Utah, University, Institution
Cluster 2, has 9 hits: 'city:8, utah:8, mormon:5, state:4, name:3, lake:3, salt
    City, Davis County, Utah
    Deseret, State, Name
    Logan, City, Seat
    Murray, City, Salt Lake County
    Nevada, State, U.S.
    Provo, City, Seat
    Salt Lake City, City, Capital
    Utah, State, U.S.
    Utah Lake, Freshwater, Lake
Cluster 3, has 3 hits: 'mormonism:3, polygamy:3, smith:3, mormon:3, church:2, ki
    Mormonism, World, Religion
    Smith, Joseph, Religious
    Brigham, Religious, Leader
Cluster 4, has 7 hits: 'utah:6, colorado:5, mi:4, km:4, river:2, yampa:2, ute:2,
    Colorado, State, United States
    Colorado, River, North America
    Salt Lake, Body, Salt
    Green, River, Utah
    Hovenweep National Monument, Colorado, Utah
    Uinta Mountains, Range, Mountain
    Ute, North American Indian, Tribe

```

CLAIMS

I claim:

1. A system for case-based organizing and querying of a database, said database having a set of objects, said system comprising

means for organizing said database, by examining each object in said database and associating that object with a first set of property values;

means responsive to a query, by associating said query with a second set of property values and performing matching on the objects of the database for objects which are similar.

2. A system as in claim 1, wherein said objects comprise text.

3. A system as in claim 1, wherein said first set of property values comprise keywords or other indicators of content.

4. A system as in claim 1, wherein said first set of property values comprise those words which appear more frequently in the document than in the database at large.

5. A system as in claim 1, wherein said first set of property values comprise those words which appear in a predetermined section of text of the object.

1           6. A system as in claim 1, wherein said first set of  
2 property values comprise those words which appear in a title of  
3 the object.

4  
5           7. A system as in claim 1, wherein said matching is  
6 case-based matching or other fuzzy associative matching.

7  
8           8. A system as in claim 1, wherein said query  
9 comprises text.

10  
11           9. A system as in claim 1, wherein said means  
12 responsive to a query associates said query with keywords or  
13 other indicators of its content.

14  
15           10. A system as in claim 1, comprising means for  
16 presenting a set of matched objects in response to said query.

17  
18           11. A system as in claim 1, comprising means  
19 responsive to refinement of said query.

20  
21           12. A system as in claim 1, comprising means  
22 responsive to iterative refinement of said query.

23  
24           13. A system as in claim 12, wherein said means  
25 responsive to iterative refinement uses a case-based technique.

26  
27  
28

1 14. A system as in claim 1, comprising means for  
2 ordering said set of matched objects in response to quality of  
3 match.

4  
5 15. A system as in claim 1, comprising means for  
6 organizing said set of matched objects.

7  
8 16. A system as in claim 15, wherein said means for  
9 organizing comprises means for grouping said set of matched  
10 objects into a set of clusters.

11  
12 17. A system as in claim 15, wherein said means for  
13 organizing comprises means for grouping said set of matched  
14 objects into a set of clusters of objects which have similar  
15 properties, which relate to similar content, which have similar  
16 likelihood to be of relevance to the query, or which have similar  
17 likelihood to be of interest to an operator posing the query.

18  
19 18. A system as in claim 15, comprising means for  
20 generating suggestions for iterative refinement of said query.

21  
22 19. A system as in claim 18, wherein said means for  
23 generating is responsive to a result of organizing matched  
24 objects.

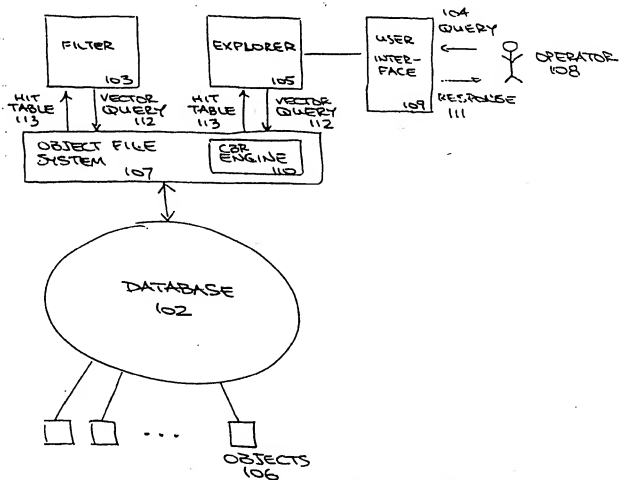
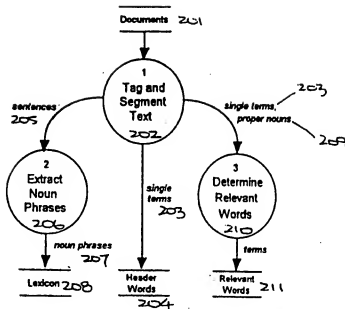
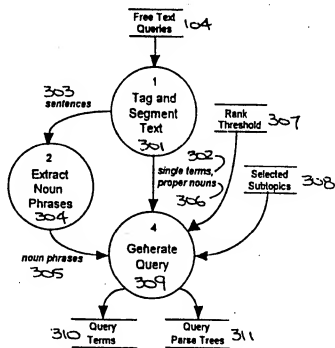


FIGURE 1.

FIGURE 2.







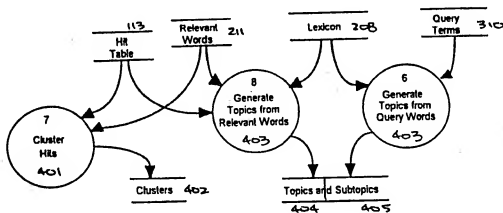


FIGURE 4.

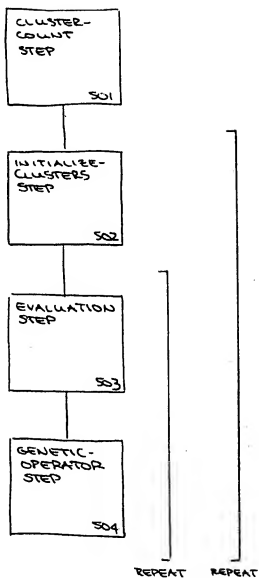
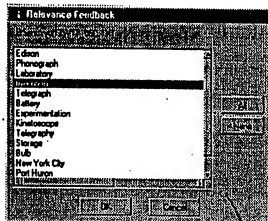
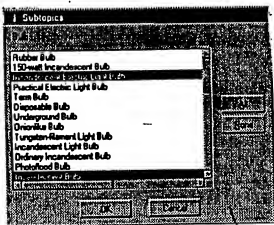
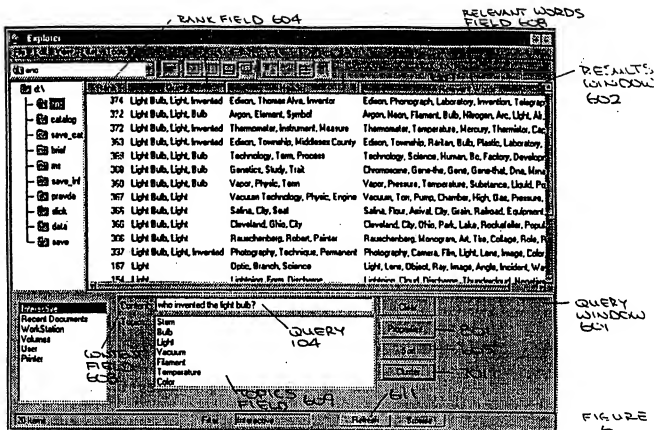


FIGURE 5.



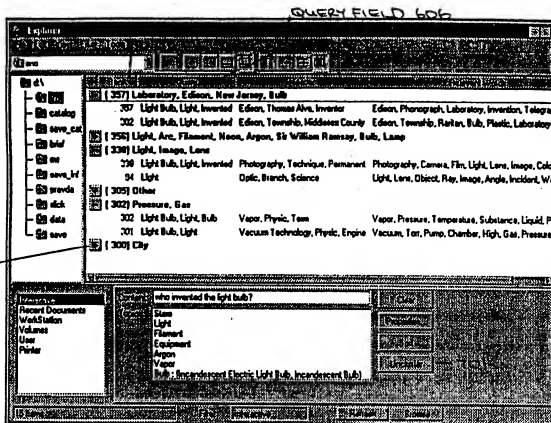


FIGURE 7.

PROPERTIES  
WINDOW  
802

**Inference Settings**

Maximum Number of Clusters (1 to 1000)	80	804
Maximum Number of Clusters (1 to 1000)	400	805
Clustering Time (in milliseconds)	2500	806
Maximum Number of Clusters (1 to 1000)	2	807
Maximum Number of Clusters (1 to 1000)	8	808
Maximum Number of Clusters (1 to 1000)	0	
Maximum Words per Secondary Level (1 to 200)	10	
Maximum Topic (1 to 20)	7	809
Maximum Subtopic (1 to 1000)	250	810
Maximum Subtopic (per Topic) (1 to 500)	80	

☒ Vector (not Boolean)  
☒ Clustering  
☒ Query Topic Generation  
☒ Query Topic Generation

FIGURE  
8.

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US94/07569

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) : G06F 15/40

US CL : 395/600; 364/419.19

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 395/600; 364/419.19

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS

search terms: information retrieval, document retrieval, case-based reasoning, cluster, keyword, index, text, full-text

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X - Y	US, A, 5,099,426 (Carlgen et al.) 24 March 1992, col. 4, lines 4-68, col. 5, lines 1-11.	1-4, 7-13 ----- 5-6, 14-19
Y, P	US, A, 5,303,361 (Colwell et al.), 12 April 1994, col. 2, lines 12-35.	14
Y	US, A, 5,062,074 (Kleinberger et al.) 29 October 1991, col. 9, lines 58-68, col. 10, lines 1-68.	15-19
X	US, A, 5,201,048 (Coulter et al.) 06 April 1993, abstract, col. 2, lines 25-68, col. 3, lines 1-48.	1-3

☐ Further documents are listed in the continuation of Box C.☐ See patent family annex.

## \* Special categories of cited documents:

\*A\* document defining the general state of the art which is not considered to be part of particular relevance

\*E\* earlier document published on or after the international filing date

\*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

\*O\* document referring to an oral disclosure, use, exhibition or other means

\*P\* document published prior to the international filing date but later than the priority date claimed

\*T\*

later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*X\*

document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\*

document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

\*Z\*

document member of the same patent family

Date of the actual completion of the international search

29 AUGUST 1994

Date of mailing of the international search report

24 OCT 1994

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Authorized officer

LARRY J. ELLCESSOR